CHAPTER 1

INTRODUCTION TO PROBABILITY AND COUNTING

What is "statistics" and why is its study important to engineers and scientists? To answer this question let us describe an aspect of the work of a scientist known as "model building."

Basically, the job of a scientist is to describe what he or she sees, to try to explain what is observed, and to use this knowledge to predict events in the world in which we live. The explanation often takes the form of a physical model. A *model* is a theoretical explanation of the phenomenon under study and, at the outset, is usually expressed verbally. To use the model for predictive purposes, this verbal description must be translated into one or more mathematical equations. These equations can be used to determine the value of a specific variable in the model based on the knowledge of the values assumed by other model variables. For example, the Perfect Gas Law states that the pressure and volume of a gas may both vary simultaneously when the temperature of the gas is changed. This verbal model can be translated into a mathematical equation by writing

$$\text{Perfect Gas Law: } PV = RT$$

where P is the pressure of the gas, V is its volume, T is its temperature, and R is a constant called the gas constant. The numerical value of the gas constant depends on the physical units chosen for the other terms in the model. Once we know the values assumed by two of the three variables P, V, or T, we can calculate the value of the third via this mathematical model. For example, under a pressure of 760 mm mercury and a temperature of 273 kelvins, a mole of any gas is thought to have a volume of 22.4 liters. The gas constant in this case has a value of approximately 62.36. Based on the Perfect Gas Law, a gas with a volume of 5 liters at a temperature of 100 kelvins has pressure P given by

$$PV = RT = 62.36T$$

or

$$P(5) = 62.36(100)$$

$$P = 1247.2 \text{ mm mercury}$$

That is, our model leads us to expect the pressure to be 1247.2 mm mercury. A model such as the Perfect Gas Law is said to be "deterministic." It is deterministic in the sense that it allows us to determine an exact value for the variable of interest under specified experimental conditions. The Perfect Gas Law does describe *some* real gases at moderate temperatures and pressures. Unfortunately, many real gases cannot be described by this or any other deterministic model especially at extreme temperatures and pressures! Under these circumstances, we must find another way to predict the behavior of the gas with some degree of certainty. This can be done with the aid of statistical methods.

What do we mean by statistical methods? These are methods by which decisions are made based on the analysis of data gathered in carefully designed experiments. Since experiments cannot be designed to account for every conceivable contingency, there is always some uncertainty in experimental science. Statistical methods are designed to *allow us to assess the degree of uncertainty present in our results*. These methods can be classed roughly into three categories: descriptive statistics, inferential statistics, and model building. By descriptive statistics, we mean those techniques, both analytic and graphical, that allow us to describe or picture a data set. Inferential statistics concerns methods by which conclusions can be drawn about a large group of objects, called the *population*, based on observing only a *sample* or a portion of the objects in the population. Model building entails the development of prediction equations from experimental data. These equations are called statistical models; they are models that allow us to predict the behavior of a complex system and to assess our probability of error. These categories are not mutually exclusive. That is, methods developed to solve problems in one area often find application in another. We shall be concerned with all three areas in this text. The mathematics upon which statistical methods rest is probability theory. For this reason, we begin the study of statistics by considering the basic concepts of probability.

1.1 INTERPRETING PROBABILITIES

When asked "Do you know anything about probability?", most people are quick to answer "no!" Usually that is not the case at all. The ability to interpret probabilities is assumed in our culture. One hears the phrases "the probability of rain today is 95%" or "there is a 0% chance of rain today." It is assumed that the general public can interpret these values correctly. The interpretation of probabilities is summarized as follows:

1. Probabilities are numbers between 0 and 1, inclusive, that reflect the chances of a physical event occurring.
2. Probabilities near 1 indicate that the event is extremely likely to occur. They mean not that the event will occur, only that the event is considered to be a common occurrence.
3. Probabilities near zero indicate that the event is not very likely to occur. They mean not that the event will fail to occur, only that the event is considered to be rare.
4. Probabilities near $1/2$ indicate that the event is just as likely to occur as not.
5. Since numbers between 0 and 1 can be expressed as percentages between 0 and 100, probabilities are expressed often as percentages. This is particularly common in writings of a nontechnical nature.

These properties are guidelines for interpreting probabilities once they are available, but they do not indicate how to assign probabilities to events. Three methods are widely used: the *personal* approach, the *relative frequency* approach, and the *classical approach*. These methods are illustrated in the following examples.

Example 1.1.1. An oil spill has occurred. An environmental scientist asks "What is the probability that this spill can be contained before it causes widespread damage to nearby beaches?" Many factors come into play, among them the type of spill, the amount of oil spilled, the wind and water conditions during the clean up operation, and the nearness of the beaches. These factors make this spill unique. The scientist is called upon to make a value judgment; to assign a probability to the event based on informed *personal opinion*.

The main advantage of the personal approach is that it is always applicable. Anyone can have a personal opinion about anything. Its main disadvantage is, of course, that its accuracy depends on the accuracy of the information available and the ability of the scientist to assess that information correctly.

Example 1.1.2. An electrical engineer is studying the peak demand at a power plant. It is observed that on 80 of the 100 days randomly selected for study from past records, the peak demand occurred between 6 and 7 p.m. It is natural to assume that the probability of this occurring on another day is at least *approximately*

$$\frac{80}{100} = .80$$

This figure is not simply a personal opinion. It is a figure based on repeated experimentation and observation. It is a *relative frequency*.

The relative frequency approach can be used whenever the experiment can be repeated many times and the results observed. In such cases, the probability of the occurrence of event A, denoted by $P[A]$, is approximated by

$$P[A] \doteq \frac{f}{n} = \frac{\text{number of times event } A \text{ occurred}}{\text{number of times experiment was run}}$$

The disadvantage in this approach is that the experiment cannot be a one-shot situation; it must be repeatable. Remember that any probability obtained this was is an approximation. It is a value based on n trials. Further testing might result in a different approximate value. However, as the number of trials increases, the changes in the approximate values obtained tend to become slight. Thus for a large number of trials, the approximate probability obtained by using the relative frequency approach is usually quite accurate.

Example 1.1.3. What is the probability that a child born to a couple heterozygous for eye color (each with genes for both brown and blue eyes) will be brown-eyed? To answer this question, we note that since the child receives one gene from each parent, the possibilities for the child are (brown, blue), (blue, brown), (blue, blue) and (brown, brown) where the first member of each pair represents the gene received from the father. Since each parent is just as likely to contribute a gene for brown eyes as for blue eyes, all four possibilities are equally likely. Since the gene for brown eyes is dominant, three of the four possibilities lead to a brown-eyed child. Hence the probability that the child is brown-eyed is $3/4 = .75$.

The above probability is not a personal opinion, nor is it based on repeated experimentation. In fact, we found this probability by the *classical* method. This method can be used *only* when it is reasonable to assume that the possible outcomes of the experiment are equally likely. In this case, the probability of the occurrence of event A is given by

$$P[A] = \frac{n(A)}{n(S)} = \frac{\text{number of ways } A \text{ can occur}}{\text{number of ways the experiment can proceed}}$$

One advantage to this method is that it does not require experimentation. Furthermore, if the outcomes are truly equally likely, then the probability assigned to event A is not an approximation. It is an accurate description of the frequency with which the event A will occur.

1.2 SAMPLE SPACES AND EVENTS

To determine what is "probable" in an experiment, we first must determine what is "possible." That is, the first step in analyzing most experiments is to make a list of possibilities for the experiment. Such a list is called a *sample space*. We define this term as follows.

Definition 1.2.1 (Sample space and sample point). A sample space for an experiment is a set S with the property that each physical outcome of the experiment corresponds to exactly one element of S. An element of S is called a sample point.

When the number of possibilities is small, an appropriate sample space usually can be found without difficulty. For instance, we have seen that when a couple heterozygous for eye color parents a child, the possible genotypes for the child are given by

$$S = \{(\text{brown}, \text{blue}), (\text{blue}, \text{brown}), (\text{blue}, \text{blue}), (\text{brown}, \text{brown})\}$$

As the number of possibilities becomes larger, it is helpful to have a system for developing a sample space. One such system is the *tree diagram*. The next example illustrates the idea.

Example 1.2.1. During a space shot, the primary computer system is backed up by two secondary systems. They operate independently of one another in that the failure of one has no effect on any of the others. We are interested in the readiness of these three systems at launch time. What is an appropriate sample space for this experiment?

Since we are primarily concerned with whether each system is operable at launch, we need only find a sample space that gives that information. To generate the sample space we use a *tree*. The primary system either is operable (yes) or not operable (no) at the time of launch. This is indicated in the tree diagram of Fig. 1.1(a), where yes = y and no = n. Likewise the first backup system either is or is not operable. This is shown in Fig. 1.1(b). Finally, the second backup system either is or is not operable. The tree is completed as shown in Fig. 1.1(c). A sample space S for the experiment can be read from the tree by following each of the eight distinct paths through the tree. Thus

$$S = \{yyy, yyn, yny, ynn, nyy, nyn, nny, nnn\}$$

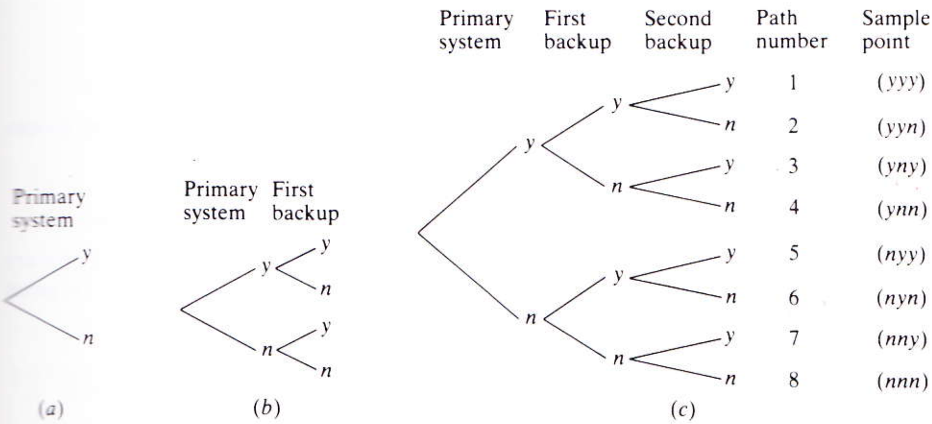


FIGURE 1.1
Constructing a tree diagram.

Once a suitable sample space has been found, elementary set theory can be used to describe physical occurrences associated with the experiment. This is done by considering what are called *events* in the mathematical sense.

Definition 1.2.2 (Event). Any subset A of a sample space is called an event. The empty set, $\varnothing$, is called the *impossible* event; the subset S is called the *certain event*.

Example 1.2.2. Consider a space shot in which a primary computer system is backed up by two secondary systems. The sample space for this experiment is

$$S = \{yyy, yyn, yny, ynn, nyy, nyn, nny, nnn\}$$

where, for example, yny denotes the fact that the primary system and second backup are operable at launch whereas the first backup is inoperable (see Example 1.2.1). Let

A: primary system is operable

B: first backup is operable

C: second backup is operable

The mathematical event corresponding to each of these physical events is found by listing the sample points that represent the occurrence of the event. Thus, we write

$$A = \{yyy, yyn, yny, ynn\}$$

$$B = \{yyy, yyn, nyy, nyn\}$$

$$C = \{yyy, yny, nyy, nny\}$$

Other events can be described using these events as building blocks. For example, the event that "the primary system *or* the first backup is operable" is given by the set $A \cup B$. Thus

$$A \cup B = \begin{array}{c}\text{primary } \textit{or} \text{ first}\\ \text{backup is operable}\end{array} = \{yyy, yyn, yny, ynn, nyy, nyn\}$$

Note that the word "or" will denote set union. The event that "the primary system *and* the first backup is operable" is given by the set $A \cap B$. That is,

$$A \cap B = \text{primary } \textit{and} \text{ first back up operable} = \{yyy, yyn\}$$

Note that the word "and" will denote set intersection. The event that "the primary system or the first backup is operable but the second backup is inoperable" is given by $(A \cup B) \cap C'$, where C' denotes the complement of set C. Thus

$$(A \cup B) \cap C' = \begin{array}{c}\text{primary } \textit{or} \text{ first backup operable}\\ \textit{but} \text{ second backup inoperable}\end{array} = \{yyn, ynn, nyn\}$$

Note that the word "but" is also translated as a set intersection; the word "not" translates as a set complement.

Let us pause briefly to consider a basic difference between the sample space

$$S_1 = \{(\text{brown}, \text{blue}), (\text{blue}, \text{brown}), (\text{blue}, \text{blue}), (\text{brown}, \text{brown})\}$$

of Example 1.1.3 and

$$S_2 = \{yyy, yyn, yny, ynn, nyy, nyn, nny, nnn\}$$

of Example 1.2.1. Since each parent is just as likely to contribute a gene for brown eyes as for blue eyes, the sample points of S_1 are equally likely. This allows us to use the classical method to find the probability that a child born to a couple heterozygous for eye color will be brown-eyed. That is, we can conclude that

$$\begin{aligned} P[A] &= P[\{(\text{brown}, \text{blue}), (\text{blue}, \text{brown}), (\text{brown}, \text{brown})\}] \\ &= \frac{n(A)}{n(S)} = \frac{3}{4} \end{aligned}$$

However, it is not correct to assume that the sample points of S_2 are equally likely. This would be true if and only if each of the three computer systems is just as likely to fail as to be operable at launch time. Our technology is much better than that! The primary question to be answered is "What is the probability that at least one system will be operable at the time of the launch?" That is, what is

$$P[\{yyy, yyn, yny, ynn, nyy, nyn, nny\}]?$$

As will be shown later, this question can be answered. However, since the sample points are not equally likely, it cannot be answered using the classical method.

Occasionally, interest centers on two or more events that cannot occur at the same time. That is, the occurrence of one event precludes the occurrence of the other. Such events are said to be *mutually exclusive*.

Example 1.2.3. Consider the sample space

$$S = \{yyy, yyn, yny, ynn, nyy, nyn, nny, nnn\}$$

of Example 1.2.1. The events

$$\begin{aligned} &A_1\text{: primary system operable} &&= \{yyy, yyn, yny, ynn\} \\ &A_2\text{: primary system inoperable} &&= \{nyy, nyn, nny, nnn\} \end{aligned}$$

are mutually exclusive. It is impossible for the primary system to be both operable and inoperable at the same time. Mathematically, A_1 and A_2 have no sample points in common. That is, $A_1 \cap A_2 = \varnothing$.

Example 1.2.3 suggests the mathematical definition of the term "mutually exclusive events."

Definition 1.2.3 (Mutually exclusive events). Two events A_1 and A_2 are mutually exclusive if and only if $A_1 \cap A_2 = \varnothing$. Events $A_1, A_2, A_3, \ldots$ are mutually exclusive if and only if $A_i \cap A_j = \varnothing$ for $i \neq j$.

1.3 PERMUTATIONS AND COMBINATIONS

As indicated in Sec. 1.1, there are several ways to determine the probability of an event. When the physical description of the experiment leads us to believe that the possible outcomes are equally likely, then we can compute the probability of the occurrence of an event using the classical method. In this case, the probability of an event A is given by

$$P[A] = \frac{n(A)}{n(S)}$$

Thus to compute a probability using the classical approach, you must be able to count two things: $n(A)$, the number of ways in which event A can occur, and $n(S)$, the number of ways in which the experiment can proceed. As the experiment becomes more complex, lists and trees become cumbersome. Alternative methods for counting must be found.

Two types of counting problems are common. The first involves *permutations* and the second, *combinations*. These terms are defined as follows:

Definition 1.3.1 (Permutation). A permutation is an arrangement of objects in a definite order.

Definition 1.3.2 (Combination). A combination is a selection of objects without regard to order.

Note that the characteristic that distinguishes a permutation from a combination is *order*. If the order in which some action is taken is important, then the problem is a permutation problem and can be solved using a technique called the multiplication principle. If order is irrelevant, then it is a combination problem and can be solved using a formula which we shall develop.

Example 1.3.1

1. Twenty different amino acids are commonly found in peptides and proteins. A pentapeptide consisting of the five amino acids

   alanine-valine-glycine-cysteine-tryptophan

   has different properties and is, in fact, a different compound from the pentapeptide

   alanine-glycine-valine-cysteine-tryptophan

   which contains the same amino acids. Peptides are permutations of amino acid units because the sequence, or order, of the amino acids in the chain is important.
2. A foundry ships engine blocks in lots of size 20. Before a lot is accepted, three blocks are selected at random and tested for hardness. Only three are tested because the testing requires that the blocks be cut in half, and is therefore

destructive. The three blocks selected constitute a combination of engine blocks. We are interested only in which three are selected; we are not interested in the order in which they are chosen.

Counting Permutations

Once a problem has been identified as being one in which order is important, the next question to be answered is: How many permutations or arrangements of the given objects are possible? This question usually can be answered by means of the *multiplication principle*.

Multiplication principle. Consider an experiment taking place in k stages. Let n_i denote the number of ways in which stage i can occur for $i = 1, 2, 3, \ldots, k$. Altogether the experiment can occur in $\prod_{i=1}^{k} n_i = n_1 \cdot n_2 \cdot n_3 \cdots n_k$ ways.

The next example illustrates the use of this principle.

Example 1.3.2. In how many ways can the five amino acids, alanine, valine, glycine, cysteine, tryphophan, be arranged to form a pentapeptide? This is a five-stage experiment since there are five amino acids which must fall into place in the chain. This is indicated by drawing five slots and mentally noting what they represent.

1st acid in chain	2d acid in chain	3d acid in chain	4th acid in chain	5th acid in chain

In how many ways can the first stage of the experiment occur? Answer: Five. There are five acids available, any one of which could fall into the first position. Indicate this by placing a 5 in the first slot.

5				
1st acid in chain	2d acid in chain	3d acid in chain	4th acid in chain	5th acid in chain

Once the first stage is complete, in how many ways can stage 2 be performed? Answer: Four. Since each pentapeptide is to contain the five amino acids mentioned, repetition of the acid first in the chain is not permitted. The second member of the chain must be one of the four acids remaining. Indicate this by placing a 4 in the second slot.

5	4			
1st acid in chain	2d acid in chain	3d acid in chain	4th acid in chain	5th acid in chain

Similar reasoning leads us to conclude that stage 3 can take place in 3 ways, stage 4 in 2 ways, and stage 5 in 1 way. By the multiplication principle there are

5	· 4	· 3	· 2	· 1	= 120
1st acid in chain	2d acid in chain	3d acid in chain	4th acid in chain	5th acid in chain	

pentapeptides that can be formed from these five amino acids.

There are several guidelines to keep in mind when using the multiplication principle:

1. Watch out for repetition versus nonrepetition. Sometimes objects can be repeated; sometimes they cannot. Whether or not repetition is allowed is determined by the physical context of the problem.
2. Watch out for subtraction. Consider event A. Occasionally it will be difficult, if not impossible, to find $n(A)$ directly. However, $S = A \cup A'$. Since A and A' have no points in common, $n(S) = n(A) + n(A')$. This implies that $n(A) = n(S) - n(A')$.
3. If there is a stage in the experiment with a special restriction, then you should think about the restriction first.

These points are illustrated in the next example.

Example 1.3.3. The DNA-RNA code is a molecular code in which the sequence of molecules provides significant genetic information. Each segment of RNA is composed of "words." Each word specifies a particular amino acid and is composed of a chain of three ribonucleotides. Each of the ribonucleotides in the chain is either adenine (A), uracil (U), guanine (G), or cytosine (C).

1. How many words can be formed? Here repetition is allowed. By the multiplication principle there are $4 \cdot 4 \cdot 4 = 64$ possible RNA words.
2. How many of these words involve some repetition? To answer this question, we use subtraction. There are 64 words possible. By the multiplication principle, $4 \cdot 3 \cdot 2 = 24$ of these have no repeated nucleotides. The remaining $64 - 24 = 40$ words must involve some repetition.
3. How many of the 64 words end with the nucleotides uracil or cytosine and have no repetition? Since there is a restriction on the last position of the chain, we consider it first by placing a 2 in the third position.

		2
1st	2d	3d

Once this restriction has been taken care of, we note that repetition is not allowed. This means that the nucleotide in position 3 cannot be used again. The first position can be filled with any of the three remaining nucleotides, and the second by either of the two that will be left at that point. By the multiplication principle, the number of words that end with uracil or cytosine and have no repetition is

$$\underset{\text{1st}}{\underline{3}} \cdot \underset{\text{2d}}{\underline{2}} \cdot \underset{\text{3d}}{\underline{2}} = 12$$

The use of the multiplication principle often results in a product of the form $n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$ where n is a positive integer. For example, we found that the number of pentapeptides that can be formed from the five different amino acids is $5\cdot 4\cdot 3\cdot 2\cdot 1$. This product can be denoted using what is called *factorial* notation.

Definition 1.3.3 (Factorial notation). Let n be a positive integer. The product $n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$ is called n factorial and is denoted by $n!$. Zero factorial, denoted by $0!$, is defined to be 1.

Using this notation, the number of pentapeptides that can be formed from five different amino acids is $5!$. Even though the need for zero factorial is not obvious yet, its purpose will become apparent soon.

One formula for counting permutations can be derived easily from the multiplication principle. Suppose that we have n distinct objects but we are going to use only r of these objects in each arrangement. How many permutations are possible in this case? Let us denote this number by ${}_nP_r$. Note that the subscript on the left denotes the number of distinct objects available, the P denotes the fact that we are counting permutations, and the subscript on the right denotes the number of objects used per arrangement. Since each permutation is to be an arrangement of r different objects, we need r slots

$$\underset{\substack{\text{1st}\\ \text{object}}}{\underline{\qquad}}\quad \underset{\substack{\text{2d}\\ \text{object}}}{\underline{\qquad}}\quad \underset{\substack{\text{3d}\\ \text{object}}}{\underline{\qquad}}\quad \cdots \quad \underset{\substack{r\text{th}\\ \text{object}}}{\underline{\qquad}}$$

Since n distinct objects are available, we have n choices for the first slot. Repetition is not allowed, so the number of permutations is given by

$$\underset{\substack{\text{1st}\\ \text{object}}}{\underline{\quad n \quad}} \cdot \underset{\substack{\text{2d}\\ \text{object}}}{\underline{(n-1)}} \cdot \underset{\substack{\text{3d}\\ \text{object}}}{\underline{(n-2)}} \cdots \underset{\substack{r\text{th}\\ \text{object}}}{\underline{\quad (?) \quad}}$$

To find the last number in the product, note that the number subtracted from n in each factor is one less than the slot number. Thus, the rth factor will be $n-(r-1)=n-r+1$. We now have that

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1)$$

Note that

$$\frac{n!}{(n-r)!} = \frac{n(n-1)(n-2)\cdots(n-r+1)\cancel{(n-r)}\cancel{(n-r-1)}\cdots\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}{\cancel{(n-r)}\cancel{(n-r-1)}\cdots\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}$$

$$= {}_nP_r$$

Substituting, we have shown that the formula for finding the number of permutations of n distinct objects taken r at a time is as stated in the next theorem.

Theorem 1.3.1. The number of permutations of n distinct objects used r at a time, denoted by ${}_nP_r$ is

$${}_nP_r = \frac{n!}{(n-r)!}$$

Example 1.3.4

1. ${}_9P_4 = \dfrac{9!}{(9-4)!} = \dfrac{9!}{5!} = \dfrac{9\cdot 8\cdot 7\cdot 6\cdot 5!}{5!} = 3024$

2. ${}_7P_7 = \dfrac{7!}{(7-7)!} = \dfrac{7!}{0!} = \dfrac{7!}{1} = 5040$

Note that to apply Theorem 1.3.1, the objects to be arranged must be distinct, no repetition is allowed, and there can be no restrictions on any position in the arrangement. This formula will not solve all your permutation problems! The multiplication principle should be the first thing that comes to mind once you realize that a problem involves order, either natural or imposed.

Counting Combinations

Thus far we have considered counting problems in which order is important. We now turn our attention to situations in which order is irrelevant. That is, we now consider problems involving combinations rather than permutations. One very useful formula for finding the number of combinations of n distinct objects selected r at a time can be derived. Note that arranging r objects taken from n that are available is a two-stage process. The r objects must first be selected; denote the number of ways to select these objects by ${}_nC_r$. The r objects selected must then be arranged in order; this can be done in $r!$ ways. By the multiplication principle, the number of arrangements of r objects taken from n is

$${}_nP_r = {}_nC_r \cdot r!$$

Solving this equation for ${}_nC_r$ and applying Theorem 1.3.1, we see that

$${}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}$$

This result is summarized in the next theorem and illustrated in Example 1.3.5.

Theorem 1.3.2. The number of combinations of n distinct objects selected r at a time, denoted by ${}_nC_r$, or $\binom{n}{r}$, is given by

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example 1.3.5

1. ${}_5C_3 = \dfrac{5!}{3!(5-3)!} = \dfrac{5!}{3!2!} = \dfrac{5 \cdot 4 \cdot 3!}{3!2 \cdot 1} = 10$

2. $\binom{5}{0} = {}_5C_0 = \dfrac{5!}{0!(5-0)!} = \dfrac{5!}{0!5!} = 1$

It is usually difficult at first to distinguish combinations from permutations. Look for the key words "select" and "arrange." The former signals that the problem involves combinations; the latter, that a permutation is sought.

Example 1.3.6. A foundry ships a lot of 20 engine blocks of which five contain internal flaws. The purchaser will select three blocks at random and test them for hardness. The lot will be accepted if no flaws are found. What is the *probability* that this lot will be accepted? To answer this question, we must count two things: the number of ways to select three engine blocks from 20, and the number of ways to select three engine blocks from 20 and obtain no flawed engines. The former quantity is given by

$${}_{20}C_3 = \frac{20!}{3!17!} = \frac{20 \cdot 19 \cdot 18 \cdot 17!}{3 \cdot 2 \cdot 1 \cdot 17!} = 1140$$

In order to obtain no flawed engines, all three of the sampled engines must be selected from among the 15 unflawed engines in the lot. This can be done in

$${}_{15}C_3 = \frac{15!}{12!3!} = \frac{15 \cdot 14 \cdot 13 \cdot 12!}{3 \cdot 2 \cdot 1 \cdot 12!} = 455$$

ways. Since the engines selected for testing are selected at random, each of the 1140 possible samples is equally likely. Using the classical approach to probability

$$P[\text{lot is accepted}] = \frac{455}{1140}$$

This section is intended only as an introduction to counting. The importance of these techniques in the study of probability will become apparent in Chap. 3.

CHAPTER SUMMARY

In this chapter we discussed how to interpret probabilities. We also presented three methods for assigning probabilities to events. These are called the personal, relative frequency, and classical approaches. We also introduced a number of important terms whose definitions you should know. These are

Sample space	Mutually exclusive events
Sample point	Permutation
Event	Combination
Impossible event	$n!$
Certain event	$0!$

In solving permutation problems we used the multiplication principle. This principle was used to derive a formula for ${}_nP_r$, the number of permutations of n distinct objects arranged r at a time. We also derived a formula for finding ${}_nC_r$, the number of combinations of n distinct objects selected r at a time.

EXERCISES

Section 1.1

1. One environmental hazard recently identified is overexposure to airborne asbestos. In a sample of 10 public buildings over 20 years old, three were found to be insulated with materials that produced an excess number of airborne asbestos bodies. What is the approximate probability that another building of this type will have this problem? What method are you using to assign this probability?
2. A government study defines a "group 1" nuclear accident to be one involving severe core damage, melting of uranium fuel, essential failure of all safety systems, and a major breach of the reactor's containment resulting in a large release of radioactivity into the atmosphere. In 1982, officials at the Nuclear Regulatory Commission estimated the probability of such an accident occurring in the United States before the year 2000 to be .02. What approach to probability do you think was used to determine this value? (*Roanoke Times*, November 1, 1982.)
3. Hemophilia is a sex-linked hereditary blood defect of males characterized by delayed clotting of the blood which makes it difficult to control bleeding even in the case of a minor injury. When a woman is a carrier of classical hemophilia, there is a 50% chance that a male child will inherit the disease. If a carrier gives birth to two sons, what is the probability that both boys will have the disease? What approach to probability are you using to answer this question?
4. The probability of having a fatal accident in the work place is assessed using the fatal accident frequency rate (FAFR). This rate is defined by

$$\text{FAFR} = \text{number of fatalities per 1000 workers during a working lifetime}$$

This rate can be viewed as giving the probability of an individual having a fatal accident while at work. What approach to probability is being used? These FAFR values were reported by I. C. Clingan ("Safety at Sea—Its Risk Management," *Interdisciplinary Science Reviews*, 1981, vol. 6, no. 1, pp. 36–48).

Occupation	**FAFR**
Metal manufacturer	8
Coal mining	12
Construction	67
Industry overall	4

What is the approximate probability that a coal miner will suffer a fatal injury? Coal mining is an industry and is taken into account when computing the industry-wide FAFR of 4. Can you explain how this rate could be so low while at least some of the components used in its computation are high?

Section 1.2

5. Fission occurs when the nucleus of an atom captures a subatomic particle called a neutron and splits into two lighter nuclei. This causes energy to be released. At the same time, other neutrons are emitted, two or three on the average. If at least one of these is captured by another fissionable nucleus, then a chain reaction is possible.

(*a*) Consider a reaction in which three neutrons are emitted initially. Let c denote that a given neutron is captured by another nucleus; let n denote that the neutron is not captured by another nucleus. Construct a tree denoting the possible behavior for these three neutrons.

(*b*) List the sample points generated by the tree.

(*c*) List the sample points that constitute each of these events:
A_1: a chain reaction is possible
A_2: all three neutrons are captured
A_3: a chain reaction is not possible

(*d*) Are A_1 and A_2 mutually exclusive?
Are A_1 and A_3 mutually exclusive?
Are A_2 and A_3 mutually exclusive?
Are A_1, A_2, A_3 mutually exclusive?

(*e*) The probability that a neutron will be captured depends on its neutron energy and is not the same for each neutron. Under these circumstances, is it correct to say that the probability that all three neutrons will be captured is $1/8$ because this can occur in only one way and there are eight paths through the tree of part (*a*)? Explain.

6. In ballistics studies conducted during World War II, it was found that, in ground-to-ground firing, artillery shells tended to fall in an elliptical pattern such as that of Fig. 1.2. The probability that a shell would fall in the inner ellipse is .50; the probability that it would fall in the outer ellipse is .95. ("Statistics and Probability Applied to Problems of Antiaircraft Fire in World War II," E. S. Pearson, *Statistics: A Guide to the Unknown*, Holden-Day, 1972, pp. 407–415.)

(*a*) A firing is considered to be a success (s) if the shell falls within the inner ellipse; otherwise, it is failure (f). Construct a tree to represent the firing of four shells in succession.

(*b*) List the sample points generated by the tree.

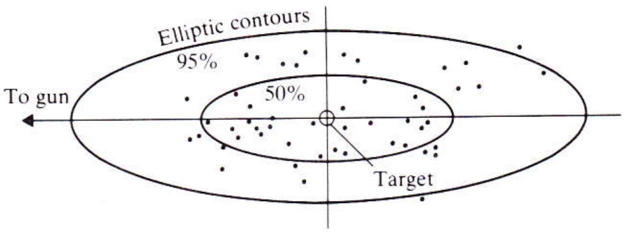


FIGURE 1.2
50% of the shells fall in the inner ellipse.

(*c*) Let A_i $i = 1, 2, 3, 4$ denote the event that the ith firing is successful. List the sample points that constitute each of the events A_1, A_2, A_3, A_4. Are these events mutually exclusive?

(*d*) List the sample points that constitute each of these events and describe the events verbally:

$$A_1'$$
$$A_1 \cup A_2$$
$$A_1 \cap A_2$$
$$A_1 \cap A_2 \cap A_3 \cap A_4$$
$$A_1 \cap A_2 \cap A_3 \cap A_4'$$
$$(A_1 \cup A_2 \cup A_3 \cup A_4)'$$
$$A_1 \cap A_1'$$

(*e*) The probability of each of the events of part (*d*) can be found using classical probability. Why is this true? Find these probabilities.

Section 1.3

7. Evaluate each of these expressions:

(*a*) 9! (*b*) 6!
(*c*) ${}_7P_3$ (*d*) ${}_6P_2$
(*e*) ${}_5P_5$ (*f*) ${}_6P_6$

8. In investigating the Ideal Gas Law, experiments are to be run at four different pressures and three different temperatures.

(*a*) How many experimental conditions are to be studied?
(*b*) If each experimental condition is replicated (repeated) five times, how many experiments will be conducted on a given gas?
(*c*) How many experiments must be conducted to obtain five replications on each experimental condition for each of six different gases?

9. In setting up a computer system for his firm to use in quality control, an engineer has four choices for the main unit: IBM, VAX, Honeywell, or HP. There are six brands of CRTs that can be purchased and three types of graphics printers.

(*a*) If all equipment is compatible, in how many ways can the system be designed?
(*b*) If the engineer wants to be able to use a statistical software package that is only available on IBM and VAX equipment, in how many ways can the system be designed?

10. In Exercise 6 we considered the experiment of firing four artillery shells in succession. Each firing was classed as being either a success or a failure. Use the multiplication rule to verify that the number of paths through the tree representing this experiment is 16.

11. The Apollo mission to land men on the moon made use of a system whose basic structure is shown in Fig. 1.3. For the system to operate successfully all five components shown must function properly. Let us identify each component as being either operable (0) or inoperable (i). Thus the sequence $0000i$ denotes a state in which all components except the LEM engine are operable. ("Striving for Reliability," Gerald Lieberman, *Statistics: A Guide to the Unknown*, Holden-Day, 1972, pp. 400–406.)
(*a*) How many states are possible?
(*b*) How many states are possible in which the LEM engine is inoperable?
(*c*) The mission is deemed at least partially successful if the first three components are operable. How many states represent at least a partially successful mission?
(*d*) The mission is a total success if and only if all five components are operable. How many states represent a completely successful mission?

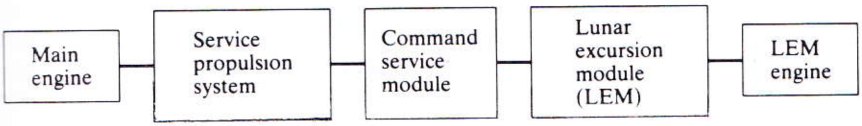


FIGURE 1.3
A simplified diagram of the Apollo system.

12. The basic storage unit of a digital computer is a "bit." A bit is a storage position that can be designated as either on (1) or off (0) at any given time. In converting picture images to a form that can be transmitted electronically, a picture element called a "pixel" is used. Each pixel is quantized into gray levels and coded using a binary code. For example, a pixel with four gray levels can be coded using two bits by designating the gray levels by 00, 01, 10, and 11.
(*a*) How many gray levels can be quantized using a four-bit code?
(*b*) How many bits are necessary to code a pixel quantized to 32 gray levels?

13. Evaluate each of these expressions:
(*a*) ${}_9C_4$ (*b*) ${}_8C_3$
(*c*) $\binom{8}{5}$ (*d*) $\binom{8}{0}$

14. Prove that ${}_nC_r = {}_nC_{n-r}$.

15. BASIC is a computer language often used on home computers. Although it is easy to learn, it is also relatively slow. A study is conducted to compare the execution time for five BASIC compilers. How many pairwise comparisons can be made among the compilers? (A study of this sort is discussed in "A Comparison of Five Compilers for Apple Basic" by J. H. and J. S. Taylor, *Byte*, 1982, vol. 7, no. 9, p. 440.)

16. The Delta project is a project to determine if large-scale agricultural production can succeed in Alaska. In this project, 22 persons are to be selected from a pool of 103 qualified applicants and awarded the right to purchase land parcels to be developed for agricultural purposes. ("Expanding Subarctic Agriculture," *Interdisciplinary Science Reviews*, 1982, vol. 7, no. 3, pp. 178–187.)

(a) In how many ways can the 22 persons be selected? (Set up only.)

(b) Assume that you are one of the persons in the applicant pool. In how many of the subgroups of part (a) will you be included? (Set up only.)

(c) If the selection process is done randomly, each of the subgroups of part (a) is equally likely. If your name is in the applicant pool, what is the probability that you will be awarded the right to purchase land?

17. (*Permutations of indistinguishable objects.*) The following formula allows us to find the number of permutations of n objects when objects of the same type are not distinct.

Consider n objects where n_1 are of type 1, n_2 of type 2, ..., n_k of type k. The number of ways in which the n objects can be arranged when objects of the same type are considered to be indistinguishable is given by

$$\frac{n!}{n_1!n_2!\cdots n_k!} \qquad n = n_1 + n_2 + \cdots + n_k$$

For example, the number of RNA words that can be formed using uracil (U) twice and guanine (G) once is $3!/2!1! = 3$.

(a) The oil embargo of 1973 spurred a study of the possibility of using automatic meter reading to reduce costs to power companies. One procedure studied entailed the use of 128-bit messages. Occasionally transmission errors occur resulting in a digit reversal of one or more bits. How many messages can be sent that contain exactly two transmission errors? *Hint:* Think of a message as being a permutation of 128 objects each of which is either correct (c) or not correct (n).

(b) In studying a chemical reaction, 12 experiments will be conducted. Four different temperatures will be used 3 times each with the temperatures run in random order. In how many orders can the series of experiments be conducted?

(c) This theorem is proved by arguing that a k-stage process is involved. Stage 1 consists of selecting n_1 positions in which to place items of type 1. Stage 2 consists of selecting n_2 positions from the $n - n_1$ that remain in which to place items of type 2. This process continues until eventually there are n_k positions remaining in which to place the items of type k. That is,

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_k}{n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$$

Verify this result for the data of part (b).

REVIEW EXERCISES

18. Find n if $\binom{n}{2} = 21$; if $\binom{n}{2} = 105$.

19. The configuration of a particular computer terminal consists of a baud-rate setting, a duplex setting, and a parity setting. There are 11 possible baud-rate settings, two parity settings (even or odd), and two duplex settings (half or full).

(a) How many configurations are possible for this terminal?

(b) In how many of these configurations is the parity even and the duplex full?

(*c*) A line surge occurs that causes these settings to change at random. What is the probability that the resulting configuration will have even parity and be full duplex?

20. A firm offers a choice of 10 free software packages to buyers of their new home computer. There are 25 packages from which to choose. In how many ways can the selection be made? Five of the packages are computer games. How many selections are possible if exactly three computer games are selected?

21. A project manager has 10 chemical engineers on her staff. Four are women and six are men. These engineers are equally qualified. In a random selection of three workers, what is the probability that no women will be selected? Would you consider it unusual for no women to be selected under these circumstances? Explain.

22. A computer system uses passwords that consist of five letters followed by a single digit.
(*a*) How many passwords are possible?
(*b*) How many passwords consist of three As and two Bs, and end in an even digit?
(*c*) If you forget your password, but remember that it has the characteristics described in part (*b*), what is the probability that you will guess the password correctly on the first attempt?

23. A mainframe computer has 16 ports. At any given time, each port is either in use or not in use. How many possibilities are there for overall port usage of this computer? How many of these entail the use of at least one port?

24. A flashlight operates on two batteries. Eight batteries are available but three are dead. In a random selection of batteries, what is the probability that exactly one dead battery will be selected?

25. An electrical control panel has three toggle switches labeled I, II, and III each of which can be either on (O) or off (F).
(*a*) Construct a tree to represent the possible configurations for these three switches.
(*b*) List the elements of the sample space generated by the tree.
(*c*) List the sample points that constitute the events
A: at least one switch is on
B: switch I is on
C: no switch is on
D: four switches are on
(*d*) Are events A and B mutually exclusive? Are events A and C mutually exclusive? Are events A and D mutually exclusive?
(*e*) What is the name given to an event such as D?
(*f*) If, at any given time, each switch is just as likely to be on as off, what is the probability that no switch is on?

26. Two items are randomly selected one at a time from an assembly line and classed as to whether they are of superior quality ($+$), average quality (0), or inferior quality ($-$).
(*a*) Construct a tree for this two-stage experiment.
(*b*) List the elements of the sample space generated by the tree.
(*c*) List the sample points that constitute the events
A: the first item selected is of inferior quality
B: the quality of each of the items is the same
C: the quality of the first item exceeds that of the second
(*d*) Are the events A and B mutually exclusive? Are the events A and C mutually exclusive?

(*e*) Give a brief verbal description of these events:
$A' \cap B$ $\quad$ $A' \cap B'$
$A \cap B'$ $\quad$ $A \cap C' \cap B$

(*f*) It is known that 90% of the items produced are of average quality, 1% are of superior quality, and the rest are of inferior quality. It is argued that since the classification experiment can proceed in nine ways with only one of these resulting in two items of average quality, the probability of obtaining two such items is 1/9. Criticize this argument.

27. An experiment consists of selecting a digit from among the digits 0 to 9 in such a way that each digit has the same chance of being selected as any other. We name the digit selected A. These lines of code are then executed.

```
IF A < 2 THEN B = 12; ELSE B = 17;
IF B = 12 THEN C = A - 1; ELSE C = 0;
```

(*a*) Construct a tree to illustrate the ways in which values can be assigned to the variables A, B, and C.
(*b*) Find the sample space generated by the tree.
(*c*) Are the 10 possible outcomes for this experiment equally likely?
(*d*) Find the probability that A is an even number.
(*e*) Find the probability that C is negative.
(*f*) Find the probability that $C = 0$.
(*g*) Find the probability that $C \leq 1$.

CHAPTER 2

SOME PROBABILITY LAWS

In Chap. 1 we considered how to interpret probabilities. In this chapter we consider some laws that govern their behavior. The laws that we shall present are those that will have a direct application to problem-solving. These laws will be stated and illustrated numerically. Their derivations are not hard and most are left as exercises.

2.1 AXIOMS OF PROBABILITY

You have probably seen the development of a mathematical system in your study of high-school geometry. In developing any mathematical system, one begins by stating a few basic definitions and axioms that underlie the system. The definitions are the technical terms of the system; axioms are statements that are assumed to be true and therefore require no proof. Usually, one starts with as few axioms as possible and then uses these axioms and the technical definitions to develop whatever theorems follow logically. Some technical terms such as sample space, sample point, event, and mutually exclusive events have already been introduced. One can develop a useful system of theorems pertaining to probability with the aid of these definitions and three axioms called the axioms of probability.

Axioms of probability.

1. Let S denote a sample space for an experiment.

$$P[S] = 1$$

2. $P[A] \geq 0$ for every event A.
3. Let $A_1, A_2, A_3, \ldots$ be a finite or an infinite sequence of mutually exclusive events. Then $P[A_1 \cup A_2 \cup A_2 \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots$.

Axiom 1 states a fact that most people regard as obvious; namely, that the probability assigned to the certain event S is 1. Axiom 2 ensures that probabilities can never be negative. Axiom 3 guarantees that when one deals with mutually exclusive events, the probability that at least one of the events will occur can be found by adding the individual probabilities. An important consequence of this axiom is that it gives us the ability to find the probability of an event when the sample points in the same space for the experiment are not equally likely. Example 2.1.1 illustrates this point.

Example 2.1.1. The distribution of blood types in the United States is roughly 41% type A, 9% type B, 4% type AB and 46% type O. An individual is brought into an emergency room and is to be blood-typed. What is the probability that the type will be A, B, or AB?

The sample space for this experiment is

$$S = \{\mathrm{A}, \mathrm{B}, \mathrm{AB}, \mathrm{O}\}$$

The sample points are not equally likely, so the classical approach to probability is not applicable. Let A_1, A_2, and A_3 denote the events that the patient has type A, B, and AB blood respectively. The events A_1, A_2, and A_3 are mutually exclusive and we are looking for $P[A_1 \cup A_2 \cup A_3]$. By axiom 3,

$$\begin{aligned} P[A_1 \cup A_2 \cup A_3] &= P[A_1] + P[A_2] + P[A_3] \\ &= .41 + .09 + .04 \\ &= .54 \end{aligned}$$

An immediate consequence of these axioms is the fact that the probability assigned to the impossible event is 0, as you should suspect. The derivation of this result is outlined in Exercise 7.

Theorem 2.1.1. $P[\varnothing] = 0$.

Another consequence of the axioms is that the probability that an event will *not* occur is equal to 1 minus the probability that it will occur. For example, if the probability of a successful space shuttle mission is .99, then the probability that it will not be successful is $1 - .99 = .01$. This idea is stated in Theorem 2.1.2. Its derivation is outlined in Exercise 8.

Theorem 2.1.2. $P[A'] = 1 - P[A]$.

The General Addition Rule

We have seen how to handle questions concerning the probability of one or another event occurring if those events are mutually exclusive. We now develop a more general rule that will allow us to find the probability that at least one of two events will occur when the events are not necessarily mutually exclusive. This rule

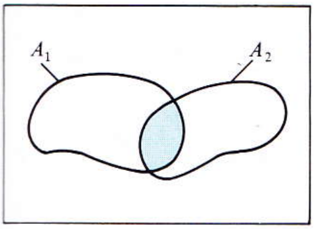


FIGURE 2.1
$A_1 \cap A_2 \neq \varnothing$.

is suggested by considering the Venn diagram of Fig. 2.1. Assume that the shaded region in the diagram, $A_1 \cap A_2$, is not empty so that A_1 and A_2 are not mutually exclusive. If we claim that

$$P[A_1 \cup A_2] = P[A_1] + P[A_2]$$

we have committed an obvious error. Since $A_1 \cap A_2 \subseteq A_1$ and $A_1 \cap A_2 \subseteq A_2$, $P[A_1 \cap A_2]$ has been included twice in our calculation. To correct this error, we subtract $P[A_1 \cap A_2]$ from the right-hand side of the equation to obtain the general addition rule

General addition rule.

$$P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$$

This rule can be derived from the axioms of probability and the theorems that we have already developed. Its proof is outlined in Exercise 11. The key word that signals its use is the word "or."

Example 2.1.2. Components of a propulsion system can be arranged in series. However, this arrangement has a serious drawback; if one component fails the system fails. This is obviously a risky arrangement for space travel! Consider a system in which the main engine has a backup. These engines are designed to operate independently in that the success or failure of one has no effect on the other. The engine component is operable if one *or* the other of these two engines is operable. Such a system is said to have the engine component in parallel. Assume that each engine is 90% reliable. That is, each functions correctly with probability .9. As we shall show later, it is then reasonable to assume that both engines operate correctly with probability .81. Find the probability that the engine component is operable. Let A_1: the main engine is operable, and A_2: the backup engine is operable. We are given that $P[A_1] = P[A_2] = .9$ and that $P[A_1 \cap A_2] = .81$. We want to find $P[A_1 \cup A_2]$. By the addition rule

$$\begin{aligned} P[A_1 \cup A_2] &= P[A_1] + P[A_2] - P[A_1 \cap A_2] \\ &= .9 + .9 - .81 = .99 \end{aligned}$$

The addition rule links the operations of union and intersection. If $P[A_1 \cap A_2]$ is known, the addition rule can be used to find $P[A_1 \cup A_2]$. Similarly, if $P[A_1 \cup A_2]$ is known, we can use the rule to find $P[A_1 \cap A_2]$. Venn diagrams are helpful when using this rule.

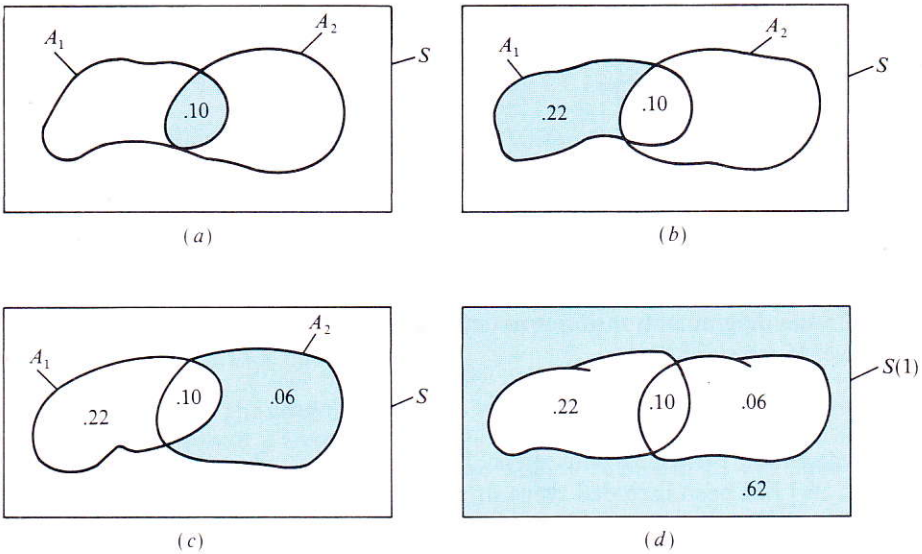


FIGURE 2.2
(*a*) $P[A_1 \cap A_2] = .10$. (*b*) $P[A_1 \cap A_2'] = .22$. (*c*) $P[A_1' \cap A_2] = .06$. (*d*) $P[A_1' \cap A_2'] = .62$.

Example 2.1.3. A chemist analyzes seawater samples for two heavy metals: lead and mercury. Past experience indicates that 38% of the samples taken from near the mouth of a river on which numerous industrial plants are located contain toxic levels of lead or mercury: 32% contain toxic levels of lead and 16% contain toxic levels of mercury. What is the probability that a randomly selected sample will contain toxic levels of lead only? Let A_1 denote the event that the sample contains toxic levels of lead, and A_2 that the sample contains toxic levels of mercury. We are given that $P[A_1] = .32$, $P[A_2] = .16$, and $P[A_1 \cup A_2] = .38$. By the addition rule

$$P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$$

or

$$.38 = .32 + .16 - P[A_1 \cap A_2]$$

Solving this equation, $P[A_1 \cap A_2] = .10$. This is indicated in Fig. 2.2(*a*). Since $P[A_1] = .32$ and $A_1 \cap A_2 \subseteq A_1$, the probability associated with the shaded region in Fig. 2.2(*b*) is .22. Similarly, since $A_1 \cap A_2 \subseteq A_2$, a probability of .06 is associated with the shaded region of Fig. 2.2(*c*). Finally since $P[S] = 1$, the probability assigned to the shaded area in Fig. 2.2(*d*) is .62. We are asked to find the probability that the sample will contain only lead. That is, we want to find $P[A_1 \cap A_2']$. This probability, .22, can be read from Fig. 2.2(*b*).

2.2 CONDITIONAL PROBABILITY

In this section we introduce the notion of conditional probability. The name itself is indicative of what is to be done. We wish to determine the probability that some event A_2 will occur, "conditional on" the assumption that some other event

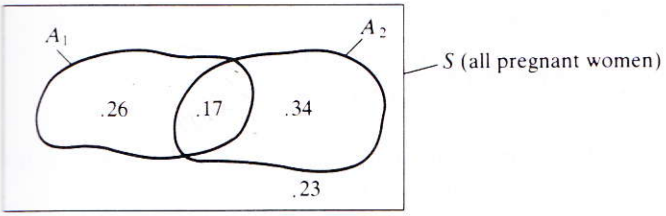


FIGURE 2.3
Partition of S.

A_1 has occurred already. The key words to look for in identifying a conditional question are "if" and "given that." We use the notation $P[A_2|A_1]$ to denote the conditional probability of event A_2 occurring given that event A_1 has occurred. A simple example will suggest the way to define this probability.

Example 2.2.1. In trying to determine the sex of a child a pregnancy test called "starch gel electrophoresis" is used. This test may reveal the presence of a protein zone called the pregnancy zone. This zone is present in 43% of all pregnant women. Furthermore, it is known that 51% of all children born are male. Seventeen percent of all children born are male and the pregnancy zone is present. The Venn diagram for these data is shown in Fig. 2.3. Let A_1 denote the event that the pregnancy zone is present, and A_2 that the child is male. We know that, for a randomly selected pregnant woman, $P[A_1] = .43$, $P[A_2] = .51$, and $P[A_1 \cap A_2] = .17$. If asked: "What is the probability that the child is male?" the answer is .51. Suppose we are *given* the information that the pregnancy zone is present and asked: "What is the probability that the child is male?" We now have information that was not available originally. What effect, if any, does this new information have on our belief that the child is male? That is, what is $P[A_2|A_1]$? Once we know that the pregnancy zone is present, our sample space no longer includes all pregnant women; it consists only of the 43% with this characteristic. Of these, $.17/.43 \doteq .395$ have male children. Logic implies that

$$P[\text{male}|\text{zone present}] = P[A_2|A_1] = .395$$

Receipt of the information that the pregnancy zone is present reduces from .51 to .395 the probability that the child is male.

To formalize the reasoning used in the previous example, note that $P[A_2|A_1]$ is found by forming a ratio whose denominator is $P[A_1]$, the probability that the *given* event will occur. The numerator is $P[A_1 \cap A_2]$, the probability that *both* the given event and the event in question will occur. That is, we define the conditional probability as follows.

Definition 2.2.1 (Conditional probability). Let A_1 and A_2 be events such that $P[A_1] \neq 0$. The conditional probability of A_2 given A_1, denoted by $P[A_2|A_1]$, is defined by

$$P[A_2|A_1] = \frac{P[A_1 \cap A_2]}{P[A_1]}$$

Sometimes receipt of the information that event A_1 has occurred has no effect on the probability assigned to event A_2. That is,

$$P[A_2|A_1] = P[A_2]$$

When this happens, A_1 and A_2 have a special relationship to one another. The nature of this relationship will be explored in the next section. In the meantime, don't be surprised if you find that a particular conditional probability does not differ from the original probability assigned to the event!

2.3 INDEPENDENCE AND THE MULTIPLICATION RULE

We have used the word "independent" informally in several previous examples. Webster's dictionary defines independent objects as objects acting "irrespective of each other." Thus two events are independent if one may occur irrespective of the other. That is, the occurrence or nonoccurrence of one does not alter the likelihood of occurrence or nonoccurrence of the other. In some cases, it is reasonable to assume that two events are independent from the physical description of the events themselves. For example, suppose that a couple heterozygous for eye color has two children. Since the eye color of a child is affected only by the genetic makeup of the parents and not by the eye color of the other child, it is reasonable to assume that the events A_1: the first child has brown eyes, and A_2: the second child has brown eyes, are independent. However, in most instances the issue is not clear-cut. In these cases we need a mathematical definition of the term to determine without a doubt whether two events are, in fact, independent.

To see how to characterize independence, let us consider a simple experiment that consists of rolling a single fair die once and then tossing a fair coin once. Let the first member of each ordered pair denote the number appearing on the die and the second, the face showing on the coin (H = heads, T = tails). A sample space for this experiment is

$$S = \{(1, H), (1, T), (2, H), (2, T), (3, H), (3, T), (4, H), (4, T), (5, H), (5, T), (6, H), (6, T)\}$$

Since the die and the coin are considered to be fair, these 12 outcomes are equally likely. Consider these events:

A: the die shows one or two

B: the coin shows heads

$A \cap B$: the die shows one or two and the coin shows heads

Since knowing the result of the die roll gives us no additional information on how the coin will land, it is reasonable to assume that the events A and B are independent. Using classical probability, it is easy to see that

$$\begin{aligned} P[A] &= P[\{(1, H), (1, T), (2, H), (2, T)\}] = 4/12 = 1/3 \\ P[B] &= P[\{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H)\}] \\ &= 6/12 = 1/2 \\ P[A \cap B] &= P[\{(1, H), (2, H)\}] = 2/12 = 1/6 \end{aligned}$$

More importantly, it is easy to see that for these physically independent events

$$P[A \cap B] = P[A] \cdot P[B]$$

Consider now an experiment that consists of drawing two coins in succession from a box containing a nickel (N), a dime (D), and a quarter (Q). The first coin is not replaced before the second is drawn. A sample space for this experiment is

$$S = \{(N, D), (N, Q), (D, N), (D, Q), (Q, N), (Q, D)\}$$

These outcomes are equally likely. Consider these events:

A: the first coin is a dime

B: the second coin is a dime

Since we do not replace the first coin before the second draw, it is evident that if event A occurs, event B cannot occur. That is, knowledge that event A has occurred does give us information on whether or not event B will occur! These events are not independent. Using classical probability, it is easy to see that

$$\begin{aligned} P[A] &= P[\{(D, N), (D, Q)\}] = 2/6 \\ P[B] &= P[\{(N, D), (Q, D)\}] = 2/6 \\ P[A \cap B] &= P[\varnothing] = 0 \end{aligned}$$

More importantly, it is easy to see that for these events that are not independent

$$P[A \cap B] \neq P[A]P[B]$$

Thus we have noticed that when A and B are clearly independent $P[A \cap B] = P[A]P[B]$; when they are clearly dependent $P[A \cap B] \neq P[A]P[B]$. This is not coincidental. It is natural to use this mathematical characterization as our technical definition of the term "independent events."

Definition 2.3.1 (Independent events). Events A_1 and A_2 are independent if and only if

$$P[A_1 \cap A_2] = P[A_1]P[A_2]$$

This definition is useful in two ways. It serves as a test for independence and it provides a way to find the probability that two events will both occur when the events are assumed to be independent. Example 2.3.1 illustrates its use as a test for independence.

Example 2.3.1. In Example 2.1.3 we considered the analysis of seawater samples taken near the mouth of a river on which numerous industrial plants are located. Let A_1 denote the event that toxic levels of lead are found, and A_2 that toxic levels of mercury are detected. We know that $P[A_1] = .32$, $P[A_2] = .16$ and $P[A_1 \cap A_2] = .10$. Are the events A_1 and A_2 independent? To decide, note that

$$P[A_1]P[A_2] = (.32)(.16) \doteq .05$$

and

$$P[A_1 \cap A_2] = .10$$

Since $P[A_1 \cap A_2] \neq P[A_1]P[A_2]$ we can conclude that these events are not independent.

illustrates the use of Definition 2.3.1 in finding the probability that two events that are assumed to be independent will occur.

Example 2.3.2. In Example 1.1.3, we found that the probability that a couple heterozygous for eye color will parent a brown-eyed child is $3/4$ for each child. Genetic studies indicate that the eye color of one child is independent of that of the other. Thus, if the couple has two children then the probability that both will be brown-eyed is

$$\begin{aligned} P\left[\begin{matrix}\text{first}\\ \text{brown}\end{matrix} \text{ and } \begin{matrix}\text{second}\\ \text{brown}\end{matrix}\right] &= P\left[\begin{matrix}\text{first}\\ \text{brown}\end{matrix}\right] P\left[\begin{matrix}\text{second}\\ \text{brown}\end{matrix}\right] \\ &= \frac{3}{4} \cdot \frac{3}{4} \\ &= \frac{9}{16} \end{aligned}$$

Definition 2.3.1 defines independence for *any* events A_1 and A_2. If at least one of the events A_1 or A_2 occurs with *nonzero* probability, then an appealing characterization of independence can be obtained. To see how this is done, assume that $P[A_1] \neq 0$. By Definition 2.3.1, A_1 and A_2 are independent if and only if

$$P[A_1 \cap A_2] = P[A_1]P[A_2]$$

Dividing by $P[A_1]$, we can conclude that A_1 and A_2 are independent if and only if

$$\frac{P[A_1 \cap A_2]}{P[A_1]} = P[A_2|A_1] = P[A_2]$$

A similar argument holds if $P[A_2] \neq 0$. We have thus derived the result given in Theorem 2.3.1

Theorem 2.3.1. Let A_1 and A_2 be events such that at least one of $P[A_1]$ or $P[A_2]$ is nonzero. A_1 and A_2 are independent if and only if

$$P[A_2|A_1] = P[A_2] \qquad \text{if } P[A_1] \neq 0$$

and

$$P[A_1|A_2] = P[A_1] \qquad \text{if } P[A_2] \neq 0$$

Since most events of real interest do occur with nonzero probability, Theorem 2.3.1 is often used as a test for independence. To understand the logic behind the theorem let us reconsider the data of Example 2.3.1.

Example 2.3.3. Consider the events A_1, a water sample contains toxic levels of lead; and A_2, a water sample contains toxic levels of mercury. We know that $P[A_1] = .32$, $P[A_2] = .16$, and $P[A_1 \cap A_2] = .10$. Suppose we are asked; "What is the probability that a randomly selected sample will contain toxic levels of mercury?" Our answer is $P[A_2] = .16$. Suppose we are now told that the sample contains toxic levels of lead and are asked: "What is the probability that toxic levels of mercury are present?" That is: "What is $P[A_2|A_1]$?" If A_1 and A_2 are independent, the new information is irrelevant and our answer should not change. That is, $P[A_2|A_1] = P[A_2]$. Otherwise, our answer should change and $P[A_2|A_1] \neq P[A_2]$. For these data, is $P[A_2|A_1] = P[A_2]$? To answer this question, note that

$$P[A_2|A_1] = \frac{P[A_1 \cap A_2]}{P[A_1]} = \frac{.10}{.32} \doteq .31$$

and

$$P[A_2] = .16$$

Since these probabilities are not the same, we conclude via Theorem 2.3.1 that A_1 and A_2 are not independent.

Occasionally, we must deal with more than two events. Again, the question arises: When are these events considered independent? Definition 2.3.2 answers this question by extending our previous definition to include more than two events.

Definition 2.3.2. Let $C = \{A_i\colon\ i = 1, 2, \ldots, n\}$ be a finite collection of events. These events are independent if and only if, given any subcollection $A_{(1)}$, $A_{(2)}, \ldots, A_{(m)}$ of elements of C

$$P\left[A_{(1)} \cap A_{(2)} \cap \cdots \cap A_{(m)}\right] = P\left[A_{(1)}\right] P\left[A_{(2)}\right] \cdots P\left[A_{(m)}\right]$$

Although this definition can be used to test a collection of events for independence, its main purpose is to provide a way to find the probability that a series of events that are assumed to be independent will occur. To illustrate, we reconsider a problem encountered in Chap. 1 (Example 1.2.1).

Example 2.3.4. During a space shot, the primary computer system is backed up by two secondary systems. They operate independently of one another and each is 90% reliable. What is the probability that all three systems will be operable at the time of the launch? Let

A_1: the main system is operable

A_2: the first backup is operable

A_3: the second backup is operable

We are given that $P[A_1] = P[A_2] = P[A_3] = .9$. We want to find $P[A_1 \cap A_2 \cap A_3]$.

Since these events are assumed to be independent

$$\begin{aligned} P[A_1 \cap A_2 \cap A_3] &= P[A_1]P[A_2]P[A_3] \\ &= (.9)(.9)(.9) \\ &= .729 \end{aligned}$$

Definition 2.3.2 must be used with care. In particular, one must be certain that it is reasonable to assume that events are independent before it is applied to compute the probability that a series of events will occur. The danger of erroneously assumed independence is illustrated in Example 2.3.5.

Example 2.3.5. An Atomic Energy Commission Study, WASH 1400, reported the probability of a nuclear accident such as that which occurred at Three Mile Island in March, 1978 to be one in 10 million. Yet, the accident did occur. According to Mark Stephens, "The methodology of WASH 1400 made use of event trees—sequences of actions that would be necessary for accidents to take place. These event trees did not assume any interrelation between events—that they might be caused by the same error in judgment or as part of the same mistaken action. The statisticians who assigned probabilities in the writing of WASH 1400 said, for example, that there was a one-in-a-thousand risk of one of the auxiliary feed-water control valves—the twelves—being closed. And if there is a one-in-a-thousand chance of one valve being closed, the chances of both valves being closed is one-thousandth of that, or a million to one. But both of the twelves were closed by the same man on March 26—and one had never been closed without the other." The events A_1: the first valve is closed, and A_2: the second valve is closed were not independent. However, they were treated as such when calculating the probability of an accident. This, among other things, led to an underestimate of the accident potential (from *Three Mile Island* by Mark Stephens, Random House, 1980).

Exercises 30 to 35 outline other interesting theorems concerning the idea of independence.

The Multiplication Rule

There is one further point to be made before we conclude this section. We can find $P[A_1 \cap A_2]$ if the events are assumed to be independent. Furthermore, if the proper information is given, the general addition rule can be used to find this probability. Is there any other way to find the probability of the simultaneous occurrence of two events if the events are not independent? The answer is yes, and the method is easy to derive. We know that

$$P[A_2|A_1] = \frac{P[A_1 \cap A_2]}{P[A_1]} \qquad P[A_1] \neq 0$$

regardless of whether the events are independent. Multiplying each side of this equation by $P[A_1]$, we obtain the following formula, called the *multiplication rule*:

Multiplication rule.

$$P[A_1 \cap A_2] = P[A_2|A_1]P[A_1]$$

The use of this rule is illustrated in Example 2.3.6.

Example 2.3.6. Recent research indicates that the approximately 49% of all infections involve anaerobic bacteria. Furthermore, 70% of all anaerobic infections are polymicrobic; that is, they involve more than one anaerobe. What is the probability that a given infection involves anaerobic bacteria *and* is polymicrobic? Let A_1 denote the event that the infection is anaerobic, and A_2 that it is polymicrobic. We are given that $P[A_1] = .49$ and that $P[A_2|A_1] = .70$. We want to find $P[A_1 \cap A_2]$. By the multiplication rule

$$\begin{aligned} P[A_1 \cap A_2] &= P[A_2|A_1]P[A_1] \\ &= (.70)(.49) \\ &= .343 \end{aligned}$$

2.4 BAYES' THEOREM (OPTIONAL)

The topic of this section is the theorem formulated by the Reverend Thomas Bayes (1761). It deals with conditional probability. Bayes' theorem is used to find $P[A|B]$ when the available information is not immediately compatible with that required to apply the definition of conditional probability directly.

Example 2.4.1 is a typical problem calling for the use of Bayes' theorem. You will find that you will apply Bayes' rule quite naturally without having seen a formal statement of the theorem!

Example 2.4.1. The blood type distribution in the United States is type A, 41%; type B, 9%; type AB, 4%; and type O, 46%. It is estimated that during World War II, 4% of inductees with type O blood were typed as having type A; 88% of those with type A were correctly typed; 4% with type B blood were typed as A; and 10% with type AB were typed as A. A soldier was wounded and brought to surgery. He was typed as having type A blood. What is the probability that this is his true blood type?

Let

A:	he has type A blood
B:	he has type B blood
AB:	he has type AB blood
O:	he has type O blood
TA:	he is typed as type A

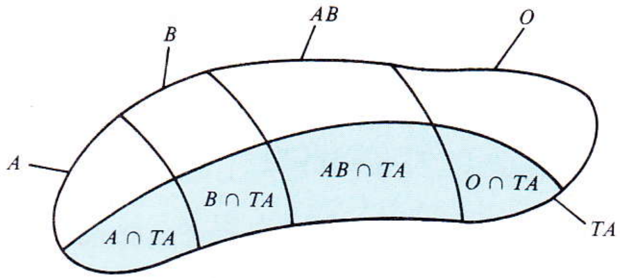


FIGURE 2.4
$TA = (A \cap TA) \cup (B \cap TA) \cup (AB \cap TA) \cup (O \cap TA)$.

We want to find $P[A|TA]$. We are given that

$$\begin{aligned} P[A] &= .41 & P[TA|A] &= .88 \\ P[B] &= .09 & P[TA|B] &= .04 \\ P[AB] &= .04 & P[TA|AB] &= .10 \\ P[O] &= .46 & P[TA|O] &= .04 \end{aligned}$$

Since the question asked is conditional, the first inclination is to try to apply the definition of conditional probability. Let us do so.

$$P[A|TA] = \frac{P[A \cap TA]}{P[TA]}$$

Unfortunately, neither $P[A \cap TA]$ nor $P[TA]$ is given. We must compute these quantities for ourselves. Each is easy to find. Note that by the multiplication rule

$$\begin{aligned} P[A \cap TA] &= P[TA|A]P[A] \\ &= (.88)(.41) \\ &\doteq .36 \end{aligned}$$

Note also that the event TA can be partitioned into four mutually exclusive events as shown in Fig. 2.4. That is,

$$TA = (A \cap TA) \cup (B \cap TA) \cup (AB \cap TA) \cup (O \cap TA)$$

By axiom 3

$$P[TA] = P[A \cap TA] + P[B \cap TA] + P[AB \cap TA] + P[O \cap TA]$$

Applying the multiplication rule to each of the terms on the right side of this equation we obtain

$$\begin{aligned} P[TA] &= P[TA|A]P[A] + P[TA|B]P[B] \\ &\quad + P[TA|AB]P[AB] + P[TA|O]P[O] \\ &= (.88)(.41) + (.04)(.09) + (.10)(.04) + (.04)(.46) \\ &\doteq .39 \end{aligned}$$

Substituting .39 for $P[TA]$ yields

$$P[A|TA] = \frac{P[A \cap TA]}{P[TA]}$$

$$= \frac{.36}{.39}$$

$$\doteq .92$$

The previous problem was solved using Bayes' rule. We now state the rule. By a partition of S we mean a collection of mutually exclusive events whose union is S.

Theorem 2.4.1 (Bayes' theorem). Let $A_1, A_2, A_3, \ldots, A_n$ be a collection of events which partition S. Let B be an event such that $P[B] \neq 0$. Then for any of the events A_j, $j = 1, 2, 3, \ldots, n$

$$P[A_j|B] = \frac{P[B|A_j]P[A_j]}{\sum_{i=1}^{n} P[B|A_i]P[A_i]}$$

To see that Bayes' theorem was used in Example 2.4.1, make these notational changes:

$$A_1 = A \qquad A_3 = AB \qquad B = TA$$

$$A_2 = B \qquad A_4 = O$$

and use the theorem to find $P[A_1|B]$. Your answer will, of course, agree with that obtained earlier.

CHAPTER SUMMARY

In this chapter we presented some of the laws that govern the behavior of probabilities. We began with the axioms and from those we were able to derive the remaining laws. In particular we derived the addition rule, which deals with the probability of the union of two events; the multiplication rule, which deals with the probability of the intersection of two events; and Bayes' theorem, which deals with conditional probability. Important terms introduced here include

Conditional probability Independent events

Care must be taken when using the concept of independence. In an applied problem, be sure that it is reasonable to assume that events A and B are independent before finding the probability of their joint occurrence via the definition $P[A \cap B] = P[A]P[B]$.

EXERCISES

Section 2.1

1. The probability that a wildcat well will produce oil is 1/13. What is the probability that it will not be productive?

2. The theft of precious metals from companies in the United States is becoming a serious problem. The estimated probability that such a theft will involve a particular metal is given below: (Based on data reported in "Materials Theft," *Materials Engineering*, February 1982, pp. 27–31.)

tin: 1/35	platinum: 1/35	nickel: 1/35
steel: 11/35	gold: 5/35	zinc: 1/35
copper: 8/35	aluminum: 2/35	silver: 4/35
titanium: 1/35		

   (Note that these events are assumed to be mutually exclusive.)
   (*a*) What is the probability that a theft of precious metal will involve gold, silver, or platinum?
   (*b*) What is the probability that a theft will not involve steel?

3. Assuming the blood type distribution to be A: 41%, B: 9%, AB: 4%, O: 46%, what is the probability that the blood of a randomly selected individual will contain the A antigen? That it will contain the B antigen? That it will contain neither the A nor the B antigen?

4. Assume that the engine component of a spacecraft consists of two engines in parallel. If the main engine is 95% reliable, the backup is 80% reliable, and the engine component as a whole is 99% reliable, what is the probability that both engines will be operable? Use a Venn diagram to find the probability that the main engine will fail but the backup will be operable. Find the probability that the backup engine will fail but the main engine will be operable. What is the probability that the engine component will fail?

5. When an individual is exposed to radiation, death may ensue. Factors affecting the outcome are the size of the dose, the length and intensity of the exposure, and the biological makeup of the individual. The term LD_{50} is used to denote the dose that is usually lethal for 50% of the individuals exposed to it. Assume that in a nuclear accident, 30% of the workers are exposed to the LD_{50} and die; 40% of the workers die; and 68% are exposed to the LD_{50} or die. What is the probability that a randomly selected worker is exposed to the LD_{50}? Use a Venn diagram to find the probability that a randomly selected worker is exposed to the LD_{50} but does not die. Find the probability that a randomly selected worker is not exposed to the LD_{50} but dies.

6. When a computer goes down, there is a 75% chance that it is due to an overload and a 15% chance that it is due to a software problem. There is an 85% chance that it is due to an overload or a software problem. What is the probability that both of these problems are at fault? What is the probability that there is a software problem but no overload?

*7. Derive Theorem 2.1.1.
   Hint: Note that $S = S \cup \varnothing$ and that S and $\varnothing$ are mutually exclusive. Apply axioms 3 and 1.

*8. Derive Theorem 2.1.2.

Hint: Note that $S = A \cup A'$ and that A and A' are mutually exclusive. Apply axioms 3 and 1.

*9. Let $A \subseteq B$. Show that $P[A] \le P[B]$.

Hint: $B = A \cup (A' \cap B)$. Apply axioms 3 and 2.

*10. Show that the probability of any event A is at most 1.

Hint: $A \subseteq S$. Apply Exercise 9 and axiom 1.

*11. Derive the addition rule.

Hint: Note that

$$A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2')$$

$$A_2 = (A_1 \cap A_2) \cup (A_1' \cap A_2)$$

$$A_1 \cup A_2 = (A_1 \cap A_2) \cup (A_1 \cap A_2') \cup (A_1' \cap A_2)$$

Apply axiom 3 to each of these expressions.

12. Let A_1 and A_2 be mutually exclusive. By axiom 3 $P[A_1 \cup A_2] = P[A_1] + P[A_2]$. Show that the general addition rule yields the same result.

Section 2.2

13. Use the data of Exercise 5 to answer these questions.
(*a*) What is the probability that a randomly selected worker will die given that he is exposed to the lethal dose of radiation?
(*b*) What is the probability that a randomly selected worker will not die given that he is exposed to the lethal dose of radiation?
(*c*) What theorem allows you to determine the answer to (*b*) from knowledge of the answer to (*a*)?
(*d*) What is the probability that a randomly selected worker will die given that he is not exposed to the lethal dose?
(*e*) Is P[die] = P[die|exposed to lethal dose]? Did you expect these to be the same? Explain.

14. Use the data of Exercise 4 to answer these questions.
(*a*) What is the probability that, in an engine system such as that described, the backup engine will function given that the main engine fails?
(*b*) Is P[backup functions] = P[backup functions|main fails]? Did you expect these to be the same? Explain.

15. In a study of waters near power plants and other industrial plants that release wastewater into the water system, it was found that 5% showed signs of chemical and thermal pollution, 40% showed signs of chemical pollution, and 35% showed evidence of thermal pollution. Assume that the results of the study accurately reflect the general situation. What is the probability that a stream that shows some thermal pollution will also show signs of chemical pollution? What is the probability that a stream showing chemical pollution will not show signs of thermal pollution?

16. A random digit generator on an electronic calculator is activated twice to simulate a random two-digit number. Theoretically, each digit from 0 to 9 is just as likely to appear on a given trial as any other digit.
(*a*) How many random two-digit numbers are possible?
(*b*) How many of these numbers begin with the digit 2?

(*c*) How many of these numbers end with the digit 9?
(*d*) How many of these numbers begin with the digit 2 and end with the digit 9?
(*e*) What is the probability that a randomly formed number ends with 9 given that it begins with a 2. Did you anticipate this result?

17. In studying the causes of power failures, these data have been gathered.

5% are due to transformer damage
80% are due to line damage
1% involve both problems

Based on these percentages, approximate the probability that a given power failure involves
(*a*) line damage given that there is transformer damage
(*b*) transformer damage given that there is line damage
(*c*) transformer damage but not line damage
(*d*) transformer damage given that there is no line damage
(*e*) transformer damage or line damage

Section 2.3

18. Let A_1 and A_2 be events such that $P[A_1] = .5$, $P[A_2] = .7$. What must $P[A_1 \cap A_2]$ equal for A_1 and A_2 to be independent?

19. Let A_1 and A_2 be events such that $P[A_1] = .6$, $P[A_2] = .4$ and $P[A_1 \cup A_2] = .8$. Are A_1 and A_2 independent?

20. Consider your answer to Exercise 13(*e*). Are the events A_1: a worker dies, and A_2: the worker is exposed to a lethal dose of radiation independent?

21. Consider your answer to Exercise 14(*b*). Are the events A_1: the backup engine functions, and A_2: the main engine fails independent?

22. Test the events A_1: a stream shows signs of thermal pollution, and A_2: a stream shows signs of chemical pollution for independence. Use the data of Exercise 15.

23. The most common water pollutants are organic. Since most organic materials are broken down by bacteria that require oxygen, an excess of organic matter may result in a depletion of available oxygen. In turn, this can be harmful to other organisms living in the water. The demand for oxygen by the bacteria is called the biological oxygen demand (BOD). A study of streams located near an industrial complex revealed that 35% have a high BOD, 10% show high acidity, and 4% have both characteristics. Are the events the stream has a high BOD and the stream has high acidity independent?

24. Studies in population genetics indicate that 39% of the available genes for determining the Rh blood factor are negative. Rh negative blood occurs if and only if the individual has two negative genes. One gene is inherited independently from each parent. What is the probability that a randomly selected individual will have Rh negative blood?

25. An individual's blood group (A, B, AB, O) is independent of the Rh classification. Find the probability that a randomly selected individual will have AB negative blood. *Hint:* See Example 2.1.1 and Exercise 24.

26. The use of plant appearance in prospecting for ore deposits is called geobotanical prospecting. One indicator of copper is a small mint with a mauve-colored flower. Suppose that, for a given region, there is a 30% chance that the soil has a high copper content and a 23% chance that the mint will be present there. If the copper content is high, there is a 70% chance that the mint will be present.

(*a*) Find the probability that the copper content will be high and the mint will be present.

(*b*) Find the probability that the copper content will be high given that the mint is present.

(c) Are the events A_1: the mint is present, and A_2: the soil has a high copper content independent?

27. A study of major flash floods that occurred over the last 15 years indicates that the probability that a flash flood warning will be issued is .5 and that the probability of dam failure during the flood is .33. The probability of dam failure given that a warning is issued is .17. Find the probability that a flash flood warning will be issued and a dam failure will occur. (Based on data reported in *McGraw-Hill Yearbook of Science and Technology*, 1980, pp. 185–186.)

***28.** The ability to observe and recall details is important in science. Unfortunately, the power of suggestion can distort memory. A study of recall is conducted as follows: Subjects are shown a film in which a car is moving along a country road. There is no barn in the film. The subjects are then asked a series of questions concerning the film. Half of the subjects are asked: "How fast was the car moving when it passed the barn?" The other half of the subjects are not asked the question. Later, each subject is asked: "Is there a barn in the film?" Of those asked the first question concerning the barn, 17% answer "yes"; only 3% of the others answer "yes." What is the probability that a randomly selected participant in this study claims to have seen the nonexistent barn? Is claiming to see the barn independent of being asked the first question about the barn? *Hint:*

$$P[\text{yes}] = P[\text{yes and asked about barn}] + P[\text{yes and not asked about barn}]$$

(Based on a study reported in *McGraw-Hill Yearbook of Science and Technology*, 1981, pp. 249–251.)

***29.** The probability that a unit of blood was donated by a paid donor is .67. If the donor was paid, the probability of contracting serum hepatitis from the unit is .0144. If the donor was not paid, this probability is .0012. A patient receives a unit of blood. What is the probability of the patient's contracting serum hepatitis from this source?

***30.** Show that the impossible event is independent of every other event.

***31.** Show that if A_1 and A_2 are independent, then A_1 and A_2' are also independent. *Hint:* $A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2')$.

***32.** Use Exercise 31 to show that if A_1 and A_2 are independent then A_1' and A_2' are also independent.

***33.** It can be shown that the result of Exercise 32 holds for any collection of n independent events. That is, if $A_1, A_2, \ldots, A_n$ are independent, then $A_1', A_2', \ldots, A_n'$ are also independent. Use this result and the data of Example 2.3.4 to find the probability that at least one of the three computer systems will be operable at the time of the launch.

***34.** Let A_1 and A_2 be mutually exclusive events such that $P[A_1]P[A_2] > 0$. Show that these events are not independent.

***35.** Let A_1 and A_2 be independent events such that $P[A_1]P[A_2] > 0$. Show that these events are not mutually exclusive.

Section 2.4

36. Use the data of Example 2.4.1 to find the probability that an inductee who was typed as having type A blood actually had type B blood.

37. A test has been developed to detect a particular type of arthritis in individuals over 50 years old. From a national survey, it is known that approximately 10% of the individuals in this age group suffer from this form of arthritis. The proposed test was given to individuals with confirmed arthritic disease, and a correct test result was obtained in 85% of the cases. When the test was administered to individuals of the same age group who were known to be free of the disease, 4% were reported to have the disease. What is the probability that an individual has this disease given that the test indicates its presence?

38. It is reported that 50% of all computer chips produced are defective. Inspection ensures that only 5% of the chips legally marketed are defective. Unfortunately, some chips are stolen before inspection. If 1% of all chips on the market are stolen, find the probability that a given chip is stolen given that it is defective.

39. As society becomes dependent upon computers, data must be communicated via public communication networks such as satellites, microwave systems, and telephones. When a message is received, it must be authenticated. This is done by using a secret enciphering key. Even though the key is secret, there is always the probability that it will fall into the wrong hands, thus allowing an unauthentic message to appear to be authentic. Assume that 95% of all messages received are authentic. Furthermore assume that only .1% of all unauthentic messages are sent using the correct key and that all authentic messages are sent using the correct key. Find the probability that a message is authentic given that the correct key is used.

REVIEW EXERCISES

40. A survey of engineering firms reveals that 80% have their own mainframe computer (M); 10% anticipate purchasing a mainframe computer in the near future (B); and 5% have a mainframe computer and anticipate buying another in the near future. Find the probability that a randomly selected firm:

(a) has a mainframe computer or anticipates purchasing one in the near future

(b) does not have a mainframe computer and does not anticipate purchasing one in the near future

(c) anticipates purchasing a mainframe computer given that it does not currently have one

(d) has a mainframe computer given that it anticipates purchasing one in the near future

Are the events M and B independent? Explain.

41. In a simulation program, three random two-digit numbers will be generated independently of one another. These numbers assume the values $00, 01, 02, \ldots, 99$ with equal probability.

(a) What is the probability that a given number will be less than 50?

(b) What is the probability that each of the three numbers generated will be less than 50?

42. A power network involves three substations A, B, C. Overloads at any of these substations might result in a blackout of the entire network. Past history has shown that if substation A alone experiences an overload then there is a 1% chance of a network blackout. For stations B and C alone these percentages are 2% and 3%. Overloads at two or more substations simultaneously result in a blackout 5% of the time. During a heat wave there is a 60% chance that substation A alone will experience an overload. For stations B and C these percentages are 20 and 15%. There is a 5% chance of an overload at two or more substations simultaneously. During a particular heat wave, a blackout due to an overload occurred. Find the probability that the overload occurred at substation A alone; substation B alone; substation C alone; two or more substations simultaneously.

43. A computer center has three printers A, B, and C which print at different speeds. Programs are routed to the first available printer. The probability that a program is routed to printers A, B, and C are .6, .3, and .1 respectively. Occasionally a printer will jam and destroy a printout. The probability that printers A, B, and C will jam are .01, .05, and .04 respectively. Your program is destroyed when a printer jams. What is the probability that printer A is involved? printer B is involved? printer C is involved?

44. A chemical engineer is in charge of a particular process at an oil refinery. Past experience indicates that 10% of all shutdowns are due to equipment failure *alone*, 5% are due to a combination of equipment failure and operator error, and 40% involve operator error. A shutdown occurs. Find the probability that

(*a*) equipment failure or operator error is involved

(*b*) operator error alone is involved

(*c*) neither operator error nor equipment failure is involved

(*d*) operator error is involved given that equipment failure occurs

(*e*) operator error is involved given that equipment failure does not occur

Are the events E: an operator error occurs, and F: an equipment failure occurs independent? Explain.

45. Assume that the probability that the air brakes on large trucks will fail on a particularly long downgrade is .001. Assume also that the emergency brakes on such trucks can stop a truck on this downgrade with probability .8. These braking systems operate independently of one another. Find the probability that

(*a*) the air brakes fail but the emergency brakes can stop the truck

(*b*) the air brakes fail and the emergency brakes cannot stop the truck

(*c*) the emergency brakes cannot stop the truck given that the air brakes fail

CHAPTER 3

DISCRETE DISTRIBUTIONS

In the sciences one often deals with "variables." Webster's dictionary defines a variable as a "quantity that may assume any one of a set of values." In statistics we deal with *random variables*—variables whose observed value is determined by chance. Many of the examples presented in previous chapters involved random variables even though the term was not used at the time. Random variables usually fall into one of two categories; they are either discrete or continuous. We begin by learning to recognize discrete random variables. The remainder of the chapter is devoted to the study of random variables of this type.

3.1 RANDOM VARIABLES

We begin by considering three examples, each of which involves a random variable. Random variables will be denoted by uppercase letters and their observed numerical values by lowercase letters.

Example 3.1.1. Consider the random variable X, the number of brown-eyed children born to a couple heterozygous for eye color. If the couple is assumed to have two children, *a priori*, before the fact, the variable X can assume any one of the values 0, 1, or 2. The variable is random in that brown eyes is dependent upon the chance inheritance of a dominant gene at conception. If, for a particular couple, there are two brown-eyed children, we write $x = 2$.

Example 3.1.2. The basic premise underlying the field of immunology is that an animal is immunized by injection of a suitable antigen. In one study, malignant plasmacytoma cells are exposed to lymphocytes carrying a specific antigen. It is hoped that these cells will fuse, since the fused cells retain the ability to grow continuously and also retain the antibody characteristics of the antigen fused. In this way the animal is quickly immunized. Cells are exposed to the lymphocytes one at a time in the presence of polyethylene glycol, a fusion-promoting agent. It is known that the probability that such a cell will fuse is $1/2$. Let Y denote the number of cells exposed to obtain the first fusion. The variable Y is random; a priori, it can assume any value in the set $\{1, 2, 3, \dots\}$. Recall from your study of calculus that a set such as this that consists of an infinite collection of isolated points is called a countably infinite set.

Example 3.1.3. In Example 1.1.2 we considered the variable T, the time at which the peak demand for electricity occurs per day. This variable is random since its value is affected by such chance factors as time of the year, humidity, and temperature. It can conceivably assume any value in the 24-hour time span from 12 midnight one day to 12 midnight the next day.

It is easy to distinguish a discrete random variable from one that is not discrete. Just ask the question: "What are the possible values for the variable?" If the answer is a finite set or a countably infinite set, then the random variable is discrete; otherwise, it is not. This idea leads to the following definition.

Definition 3.1.1 (Discrete random variable). A random variable is discrete if it can assume at most a finite or a countably infinite number of possible values.

The random variable X, the number of brown-eyed children in a two-child family, is discrete. Its set of possible values is the finite set $\{0, 1, 2\}$. The set $\{1, 2, 3, \dots\}$ of possible values for Y, the number of cells exposed to obtain the first fusion of Example 3.1.2, is countably infinite. Thus, Y is also a discrete random variable. The random variable T, the time of the peak demand for electricity at a power plant, is different from the others. Time is measured continuously and T can conceivably assume any value in the interval $[0, 24)$ where 0 denotes 12 midnight one day and 24 denotes 12 midnight the next. This set of real numbers is neither finite nor countably infinite. Any time that you ask yourself the question: "What are the possible values for the random variable?", and are forced to admit that the set of possibilities includes some interval or continuous span of real numbers, then the random variable being studied is not discrete.

3.2 DISCRETE PROBABILITY DENSITIES

When dealing with a random variable, it is not enough just to determine what values are possible. We also need to determine what is probable. We need to be able to predict in some sense the values that the variable is likely to assume at any time. Since the behavior of a random variable is governed by chance, these predictions must be made in the face of a great deal of uncertainty. The best that can be done is to describe the behavior of the random variable in terms of probabilities. Two functions are used to accomplish this. We shall refer to these as the *density function* and the *cumulative distribution function*. The former is known by a variety of names in the discrete case. Some of the most commonly encountered ones being the probability function, the probability mass function, and the probability density function. In the discrete case, the density is denoted by either $p(x)$ or $f(x)$; in the continuous case it is almost always denoted by $f(x)$. For consistency, we shall use $f(x)$ for the density in both cases. We begin by defining the density function for discrete random variables.

Definition 3.2.1 (Discrete density). Let X be a discrete random variable. The function f given by

$$f(x) = P[X = x]$$

for x real is called the density function for X.

There are several things to note concerning the density in the discrete case. First, f is defined on the entire real line, and for any given real number x, $f(x)$ is the probability that the random variable X assumes the value x. For example, $f(2)$ is the probability that the random variable X assumes the numerical value of 2. Second, since $f(x)$ is a probability, $f(x) \geq 0$ regardless of the value of x. Third, if we sum f over all values of X that occur with nonzero probability, the sum must be 1. In fact, the two conditions

1. $f(x) \geq 0$
2. $\sum_{\text{all } x} f(x) = 1$

are necessary and sufficient conditions for a function f to be a discrete density. The next example illustrates these ideas.

Example 3.2.1. Consider the random variable Y, the number of cells exposed to antigen-carrying lymphocytes in the presence of polyethylene glycol to obtain the first fusion (see Example 3.1.2). We know that under these conditions the probability that a given cell will fuse is $1/2$. Thus, the probability that it will not fuse is also $1/2$. It is reasonable to assume that the cells behave independently. The possible values for Y are $\{1, 2, 3, \ldots\}$. The probability that the first cell will fuse is $1/2$. That is,

$$P[Y = 1] = f(1) = 1/2$$

The probability that the first cell will not fuse but the second one will, yielding a value of 2 for Y, is

$$\begin{aligned} P[Y = 2] = f(2) &= P[\text{first cell does not fuse}]\,P[\text{second cell does fuse}] \\ &= 1/2 \cdot 1/2 = 1/4 \end{aligned}$$

Similarly

$$P[Y = 3] = f(3) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$$

We can summarize the entire probability structure for Y in a density table. This is a table giving the possible values for the random variable in the first row and their corresponding probabilities in the second. Note that there is an obvious pattern to the entries in row 2. When this occurs, we can find a closed form

TABLE 3.1

y	1	2	3	4 ...
$P[Y = y] = f(y)$	$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$

expression for the density. In this case

$$f(y) = \begin{cases} (1/2)^y & y = 1, 2, 3, \ldots \\ 0 & \text{elsewhere} \end{cases}$$

Is this really a density? This function is obviously nonnegative but does it sum to 1? To see, note that

$$\sum_{\text{all } y} f(y) = \sum_{y=1}^{\infty} (1/2)^y$$

is a geometric series with first term $a = 1/2$ and common ratio $r = 1/2$. From elementary calculus, such a series is known to sum to $a/(1 - r)$ provided $|r| < 1$. Thus

$$\sum_{y=1}^{\infty} (1/2)^y = \frac{a}{1 - r} = \frac{1/2}{1 - 1/2} = 1$$

and the function f is a density.

Even though a discrete density is defined on the entire real line, it is only necessary to specify the density for those values y for which $f(y) \neq 0$. For instance, in the previous example we can write

$$f(y) = (1/2)^y \qquad y = 1, 2, 3, \ldots$$

It is understood that $f(y) = 0$ for all other real numbers.

Once it is known that a function is a density, it can be used to answer questions concerning the behavior of Y.

Example 3.2.2. What is the probability that we will need to expose four or more cells to antigen-carrying lymphocytes in the presence of polyethylene glycol to obtain the first fusion? That is: What is $P[Y \geq 4]$? The density for Y is

$$f(y) = (1/2)^y \qquad y = 1, 2, 3, \ldots$$

Although the desired probability can be found directly, it is easier to use subtraction.

$$\begin{aligned} P[Y \geq 4] &= 1 - P[Y < 4] \\ &= 1 - P[Y \leq 3] \\ &= 1 - (P[Y = 1] + P[Y = 2] + P[Y = 3]) \\ &= 1 - (f(1) + f(2) + f(3)) \\ &= 1 - \left((1/2)^1 + (1/2)^2 + (1/2)^3\right) \\ &= 1 - (1/2 + 1/4 + 1/8) \\ &= 1 - 7/8 = 1/8 \end{aligned}$$

TABLE 3.2

y	1	2	3	4 ...
$P[Y \leq y] = F(y)$	$\frac{8}{16}$	$\frac{12}{16}$	$\frac{14}{16}$	$\frac{15}{16}$

The second function used to compute probabilities is the cumulative distribution function F. Most of the statistical tables used in the material that follows are tables of the cumulative distribution function for some pertinent random variable.

Definition 3.2.2 (Cumulative distribution—discrete). Let X be a discrete random variable with density f. The cumulative distribution function for X, denoted by F, is defined by

$$F(x) = P[X \leq x] \qquad \text{for } x \text{ real}$$

Consider a specific real number x_0. To find $P[X \leq x_0] = F(x_0)$, we sum the density f over all values of X that occur with nonzero probability that are less than or equal to x_0. That is, computationally,

$$F(x_0) = \sum_{x \leq x_0} f(x)$$

This idea is illustrated in Example 3.2.3.

Example 3.2.3. Consider the random variable Y of Example 3.2.1 with density

$$f(y) = (1/2)^y \qquad y = 1, 2, 3, \ldots$$

A partial cumulative table for Y is shown in Table 3.2. It is formed by summing the probabilities given in the density table, Table 3.1. It is helpful to have a closed form expression for F. In this case, it is easy to obtain such an expression. By definition,

$$F(y_0) = \sum_{y \leq y_0} f(y)$$

If we let $[y_0]$ denote the greatest integer less than or equal to y_0, then in this case $F(y_0)$ can be expressed as

$$F(y_0) = \sum_{y=1}^{[y_0]} (1/2)^y$$
$$= \sum_{y=1}^{[y_0]} (1/2)(1/2)^{y-1}$$

Recall from elementary calculus that the sum of the first n terms of a geometric series is given by

$$\sum_{k=1}^{n} ar^{k-1} = \frac{a(1 - r^n)}{1 - r} \qquad r \neq 1$$

where a is the first term of the series and r is the common ratio. Applying this result with $a = 1/2$ and $r = 1/2$,

$$F(y_0) = \frac{1/2\left[1 - (1/2)^{[y_0]}\right]}{1 - 1/2}$$
$$= 1 - (1/2)^{[y_0]}$$

The probability that at most seven cells must be exposed to obtain the first fusion is given by

$$P[Y \leq 7] = F(y) = 1 - (1/2)^7 = \frac{127}{128}$$

Some of the mathematical properties of discrete distribution functions are outlined in Exercises 11, 12, and 13.

3.3 EXPECTATION AND DISTRIBUTION PARAMETERS

The density function of a random variable completely describes the behavior of the variable. However, associated with any random variable are constants, or "parameters," that are descriptive. Knowledge of the numerical values of these parameters gives the researcher quick insight into the nature of the variables. We consider three such parameters: the mean μ, the variance σ^2, and the standard deviation σ. If the exact density of the random variable is known, then the numerical value of each parameter can be found from mathematical considerations. That is the topic of this section. If the only thing available to the researcher is a set of observations on the random variable (a data set), then the values of these parameters cannot be found exactly. They must be approximated by using statistical techniques. That is the topic of much of the remainder of this text.

To understand the reasoning behind most statistical methods, it is necessary to become familiar with one general concept, namely, the idea of *mathematical expectation* or *expected value*. This concept is used in defining many statistical parameters and provides the logical basis for most of the methods of statistical inference presented later in this text.

Definition 3.3.1 (Expected value). Let X be a discrete random variable with density f. Let $H(X)$ be a random variable. The expected value of $H(X)$, denoted by $E[H(X)]$, is given by

$$E[H(X)] = \sum_{\text{all } x} H(x)f(x)$$

provided $\sum_{\text{all } x} |H(x)| f(x)$ is finite. Summation is over all values of X that occur with nonzero probability.

There are three things to note concerning this definition. First, $H(X)$ denotes a function of X. We shall be interested in functions such as $H(X) = X$,

$H(X) = X^2$, $H(X) = (X - c)^2$ where c is a constant, and $H(X) = e^{tX}$ as these functions are especially useful in statistical theory. Second, the restriction that $\sum_{\text{all } x}|H(x)|f(x)$ exists is not particularly restrictive in practice. If the set of possible values for X is finite, it *will* be satisfied; if the set of possible values for X is countably infinite it will *usually* be satisfied. However, it is possible to concoct a density f and a function $H(X)$ for which the series $\sum_{\text{all } x}|H(x)|f(x)$ does not converge. (See Exercise 22.) In this case we say that the expected value of the random variable $H(X)$ does not exist. Third, the expected value of the random variable gives us the *long run theoretical average value* for the variable. This point is illustrated in Example 3.3.1. Please realize that the density has been greatly oversimplified for purposes of illustration!

Example 3.3.1. A drug is used to maintain a steady heart rate in patients who have suffered a mild heart attack. Let X denote the number of heartbeats per minute obtained per patient. Consider the hypothetical density given in Table 3.3. What is the average heart rate obtained by all patients receiving this drug? That is, What is $E[X]$? By Definition 3.3.1

$$\begin{aligned} E[X] &= \sum_{\text{all } x} H(x)f(x) \\ &= \sum_{\text{all } x} xf(x) \\ &= 40(.01) + 60(.04) + 68(.05) + \cdots + 100(.01) \\ &= 70 \end{aligned}$$

Since the number of possible values for X is finite, $\sum_{\text{all } x}|x|f(x)$ exists. Thus, we can say that the *average* heart rate obtained by patients using this drug is 70 heartbeats per minute. Intuitively, we should have expected this result. Notice the symmetry of the density. In the long run, we would expect as many patients with heart rates of 100 as with heart rates of 40; as many with a rate of 60 as with a rate of 80. Similarly, the rates of 68 and 72 occur with the same frequency. Each of these pairs averages to 70, the value obtained by the remaining 80% of the patients. Common sense points to 70 as the expected value for X.

When used in a statistical context, the expected value of a random variable X is referred to as its *mean* and is denoted by μ or μ_X. That is, the terms *expected value* and *mean* are interchangeable, as are the symbols $E[X]$ and μ. The mean can be thought of as a measure of the "center of location" in the sense that it indicates where the "center" of the density lies. For this reason, the mean is often referred to as a "location" parameter.

TABLE 3.3

x	40	60	68	70	72	80	100
$f(x)$	.01	.04	.05	.80	.05	.04	.01

There are three rules for handling expected values that are useful in justifying statistical procedures in later chapters. These rules hold for both continuous and discrete random variables. The rules are stated and illustrated here. We outline the proofs of the first two as exercises; the proof of rule 3 must be deferred until Chap. 5.

Theorem 3.3.1 (Rules for expectation). Let X and Y be random variables and let c be any real number.

1. $E[c] = c$ (The expected value of any constant is that constant.)
2. $E[cX] = cE[X]$ (Constants can be factored from expectations.)
3. $E[X + Y] = E[X] + E[Y]$ (The expected value of a sum is equal to the sum of the expected values.)

Example 3.3.2. Let X and Y be random variables with $E[X] = 7$ and $E[Y] = -5$. Then

$$\begin{aligned} E[4X - 2Y + 6] &= E[4X] + E[-2Y] + E[6] && \text{Rule 3} \\ &= 4E[X] + (-2)E[Y] + E[6] && \text{Rule 2} \\ &= 4E[X] - 2E[Y] + 6 && \text{Rule 1} \\ &= 4(7) - 2(-5) + 6 \\ &= 44 \end{aligned}$$

Knowledge of the mean of a random variable is important but this knowledge *alone* can be misleading. The next example should show you the problem.

Example 3.3.3. Suppose that we wish to compare a new drug to that of Example 3.3.1. Let X denote the number of heartbeats per minute obtained using the old drug and Y the number per minute obtained with the new. The hypothetical density of each of these variables is given in Table 3.4. Since each of the densities is symmetric, inspection shows that $\mu_X = \mu_Y = 70$. Each drug produces *on the average* the same number of heartbeats per minute. However, there is obviously a drastic difference between the two drugs that is not being detected by the mean. The old drug produces fairly consistent reactions in patients, with 90% differing from the mean by at most 2; very few (2%) have an extreme reaction to the drug. However, the new drug produces highly diverse responses. Only 10% of the patients have heart

TABLE 3.4

x	40	60	68	70	72	80	100
$f(x)$	.01	.04	.05	.80	.05	.04	.01

y	40	60	68	70	72	80	100
$f(y)$	.40	.05	.04	.02	.04	.05	.40

rates within 2 units of the mean, whereas 80% show an extreme reaction. If we examined only the mean, we would conclude that the two drugs had identical effects —but nothing could be further from the truth!

It is obvious from Example 3.3.3 that something is not being measured by the mean. That something is *variability*. We must find a parameter that reflects consistency or the lack of it. We want the measure to assume a large positive value if the random variable fluctuates in the sense that it often assumes values far from its mean; the measure should assume a small positive value if the values of X tend to cluster closely about the mean. There are several ways to define such a measure. The most widely used is the *variance*.

Definition 3.3.2 (Variance). Let X be a random variable with mean μ. The variance of X, denoted by Var X, or σ^2, is given by

$$\operatorname{Var} X = \sigma^2 = E\left[(X - \mu)^2\right]$$

Note that the variance measures variability by considering $X - \mu$, the difference between the variable and its mean. The difference is squared so that negative values will not cancel positive ones in the process of finding the expected value. When expressed in the form $E[(X - \mu)^2]$, it is easy to see that σ^2 has the properties that we want. When the variable X often assumes values far from μ, σ^2 will be a large positive number; when the values of X tend to fall close to μ, σ^2 will assume a small positive value. Usually, the definition of σ^2 is not used to compute the variance. Rather, we use an alternative form which is given in the following theorem.

Theorem 3.3.2 (Computational formula for σ^2)

$$\sigma^2 = \operatorname{Var} X = E[X^2] - (E[X])^2$$

Proof. By definition

$$\begin{aligned}\operatorname{Var} X &= E\left[(X - \mu)^2\right]\\ &= E[X^2 - 2\mu X + \mu^2]\end{aligned}$$

Using the rules of expectation, Theorem 3.3.1,

$$\operatorname{Var} X = E[X^2] - 2\mu E[X] + \mu^2.$$

Since the symbols μ and $E[X]$ are interchangeable

$$\begin{aligned}\operatorname{Var} X &= E[X^2] - 2(E[X])^2 + (E[X])^2\\ &= E[X^2] - (E[X])^2\end{aligned}$$

We illustrate the theorem by computing the variance of each of the random variables of Example 3.3.3.

Example 3.3.4. To find σ_X^2 and σ_Y^2 for the variables of Example 3.3.3, we first use Table 3.4 to find $E[X^2]$ and $E[Y^2]$. We know that $E[X] = E[Y] = 70$.

$$
\begin{aligned}
E[X^2] &= \sum_{\text{all } x} x^2 f(x) \\
&= (40^2)(.01) + (60^2)(.04) + \cdots + (100^2)(.01) \\
&= 4926.4 \\
E[Y^2] &= \sum_{\text{all } y} y^2 f(y) \\
&= (40^2)(.40) + (60^2)(.05) + \cdots + (100^2)(.40) \\
&= 5630.32
\end{aligned}
$$

By Theorem 3.3.2

$$
\begin{aligned}
\operatorname{Var} X &= E[X^2] - (E[X])^2 \\
&= 4926.4 - 70^2 = 26.4 \\
\operatorname{Var} Y &= E[Y^2] - (E[Y])^2 \\
&= 5630.32 - 70^2 = 730.32
\end{aligned}
$$

As expected, $\operatorname{Var} Y > \operatorname{Var} X$. Even though the drugs produce the same mean number of heartbeats per minute, they do not behave in the same way. The new drug is not as consistent in its effect as the old.

Note that the variance of a random variable reported alone is not very informative. Is a variance of 26.4 large or small? Only when this value is compared to the variance of a similar variable does it take on meaning. Hence variances are used often for comparative purposes to choose between two variables which otherwise appear to be identical. Also note that the variance of a random variable is essentially a pure number whose associated units are often physically meaningless. When this occurs the unit can be omitted. For example, the unit associated with the variance of Example 3.3.4 is a "squared heartbeat." This makes little sense, so in this case variance can be reported with no unit attached. To overcome this problem, a second measure of variability is employed. This measure is the nonnegative square root of the variance, and it is called the *standard deviation*. It has the advantage of having associated with it the same units as the original data.

Definition 3.3.3 (Standard deviation). Let X be a random variable with variance σ^2. The standard deviation of X, denoted by σ, is given by

$$\sigma = \sqrt{\operatorname{Var} X} = \sqrt{\sigma^2}$$

Example 3.3.5. The standard deviations of variables X and Y of Example 3.3.4 are, respectively,

$$
\begin{aligned}
\sigma_X &= \sqrt{\operatorname{Var} X} = \sqrt{26.4} = 5.14 \text{ heartbeats per minute} \\
\sigma_Y &= \sqrt{\operatorname{Var} Y} = \sqrt{730.32} = 27.02 \text{ heartbeats per minute}
\end{aligned}
$$

Just as there are three rules for expectation, that help in simplifying complex expressions, there are three rules for variance. These rules parallel those for expectation. Rules 1 and 2 can be proved using the rules for expectation (see Exercise 20). The proof of rule 3 must be deferred until the notion of "independent random variables" has been formalized.

Theorem 3.3.3 (Rules for variance). Let X and Y be random variables and c any real number. Then

1. $\operatorname{Var} c = 0$
2. $\operatorname{Var} cX = c^2 \operatorname{Var} X$
3. If X and Y are independent, then $\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y$

(Two variables are independent if knowledge of the value assumed by one gives no clue to the value assumed by the other.)

Example 3.3.6. Let X and Y be independent with $\sigma_X^2 = 9$ and $\sigma_Y^2 = 3$. Then

$$\begin{aligned}
\operatorname{Var}[4X - 2Y + 6] &= \operatorname{Var}[4X] + \operatorname{Var}[-2Y] + \operatorname{Var} 6 && \text{Rule 3} \\
&= 16 \operatorname{Var} X + 4 \operatorname{Var} Y + \operatorname{Var} 6 && \text{Rule 2} \\
&= 16 \operatorname{Var} X + 4 \operatorname{Var} Y + 0 && \text{Rule 1} \\
&= 16(9) + 4(3) = 156
\end{aligned}$$

In this section we discussed three *theoretical* parameters associated with a random variable X. We showed not only how to determine their numerical values from knowledge of the density, but also how to interpret them physically. Keep these things in mind, for they play a major role in the study of statistical methods for analyzing experimental data.

3.4 MOMENT GENERATING FUNCTION AND THE GEOMETRIC DISTRIBUTION

Thus far, we have considered properties common to all discrete random variables. We now turn our attention to the discussion of some particular types of discrete random variables that arise in the physical world. The variables form "families" in the sense that each member of the family is characterized by a density function of the same mathematical form, differing only with respect to the numerical value of some pertinent parameter(s).

Geometric Distribution

We begin by considering the family of *geometric* random variables. As you shall see, you have already encountered some random variables of this type even though the name "geometric random variable" was not mentioned at the time.

Geometric random variables arise in practice in experiments characterized by these properties:

Geometric properties

1. The experiment consists of a series of trials. The outcome of each trial can be classed as being either a "success" (s) or a "failure" (f). A trial with this property is called a *Bernoulli* trial.
2. The trials are identical and independent in the sense that the outcome of one trial has no effect on the outcome of any other. The probability of success, p, remains the same from trial to trial.
3. The random variable X denotes the number of trials needed to obtain the first success.

The sample space for an experiment such as that just described is

$$S = \{s, fs, ffs, fffs, \dots\}$$

Since the random variable X denotes the number of trials needed to obtain the first success, X assumes the values $1, 2, 3, 4, \dots$. To find the density for X we look for a pattern. Note that

$$P[X = 1] = P[\text{success on first trial}] = p$$

$$P[X = 2] = P[\text{fail on first trial and succeed on second trial}]$$

Since the trials are independent, the latter probability can be found by multiplying. That is,

$$\begin{aligned} P[X = 2] &= P[\text{fail on first trial and succeed on second trial}] \\ &= P[\text{fail on first trial}]P[\text{succeed on second trial}] \\ &= (1 - p)(p) \end{aligned}$$

Similarly

$$\begin{aligned} P[X = 3] &= P[\text{fail on first trial and fail on second trial and succeed on third trial}] \\ &= (1 - p)(1 - p)(p) = (1 - p)^2 p \end{aligned}$$

You should be able to see that the density for X is given by Table 3.5. As you can see, the probabilities given in row 2 of the table exhibit a definite pattern. This

TABLE 3.5

x	1	2	3	4	5	$\cdots$
$f(x)$	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4p$	$\cdots$

pattern can be expressed in closed form as

$$f(x) = (1-p)^{x-1}p \qquad x = 1, 2, 3, \ldots$$

We now define a geometric random variable as being any random variable with a density of this form.

Definition 3.4.1 (Geometric distribution). A random variable X is said to have a geometric distribution with parameter p if its density f is given by

$$f(x) = (1-p)^{x-1}p \qquad 0 < p < 1$$
$$x = 1, 2, 3, \ldots$$

The function f given in this definition is a density. It is obviously nonnegative. Furthermore,

$$\sum_{x=1}^{\infty} (1-p)^{x-1}p$$

is a geometric series with first term $a = p$ and common ratio $r = (1-p)$. Thus, the series sums to

$$\frac{a}{1-r} = \frac{p}{1-(1-p)} = 1$$

as desired. From this argument, the reason for the name "geometric" distribution should be apparent.

Example 3.4.1. Random digits are integers selected from among $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ one at a time in such a way that at each stage in the selection process the integer chosen is just as likely to be one digit as any other. In simulation experiments it is often necessary to generate a series of random digits. This can be done in a number of ways, the most common being by means of a computerized random number generator. In generating such a series, let X denote the number of trials needed to obtain the first zero. This experiment consists of a series of independent, identical trials with "success" being the generation of a zero. The probability of success is $p = 1/10$. Since X denotes the number of trials needed to obtain the first success, X is a geometric random variable. Its density is found by substituting the value $1/10$ for p in the expression for f given in Definition 3.4.1. That is,

$$f(x) = (1-p)^{x-1}p \qquad x = 1, 2, 3, \ldots$$

or

$$f(x) = (9/10)^{x-1}1/10 \qquad x = 1, 2, 3, \ldots$$

Finding the mean of a geometric random variable from the definition is tricky! Consider the next example.

Example 3.4.2. Let us find the mean of the random variable X, the number of trials needed to obtain a zero when generating a series of random digits.

By Definition 3.3.1

$$\mu = E[X] = \sum_{x=1}^{\infty} xf(x)$$

$$= \sum_{x=1}^{\infty} x(9/10)^{x-1}1/10$$

That is,

$$E[X] = 1/10 + 18/100 + 243/1000 + 2916/10{,}000 + \cdots$$

This series is not geometric. Consider the series $(9/10)E[X]$.

$$(9/10)\,E[X] = 9/100 + 162/1000 + 2187/10{,}000 + 26{,}244/100{,}000 + \cdots$$

Subtracting the latter from the former, we obtain

$$(1/10)\,E[X] = 1/10 + 9/100 + 81/1000 + 729/10{,}000 + \cdots$$

This series is geometric with first term $1/10$ and common ratio $9/10$. Thus

$$(1/10)\,E[X] = \frac{1/10}{1 - 9/10} = 1$$

or

$$E[X] = \frac{1}{1/10} = 10$$

Moment Generating Function

As we have seen, the two expectations $E[X]$ and $E[X^2]$ are very useful as they allow us to find the mean and variance of the random variable. These, and other expectations of the form $E[X^k]$ for k a positive integer are called *ordinary moments*. Thus, $E[X] = \mu$ is the first ordinary moment for X; $E[X^2]$ is its second ordinary moment. The preceding example shows that finding ordinary moments, even the first moment, from the definition of expectation is not always easy. Fortunately, it is often possible to obtain a function, called the *moment generating function*, which will enable us to find these moments with less effort.

Definition 3.4.2 (Moment generating function). Let X be a random variable with density f. The moment generating function for X (m.g.f.) is denoted by $m_X(t)$ and is given by

$$m_X(t) = E[e^{tX}]$$

provided this expectation is finite for all real numbers t in some open interval $(-h, h)$.

Since each geometric random variable has a density of the same general form, it is possible to find a general expression for the moment generating function for such a variable. This expression is given in Theorem 3.4.1.

Theorem 3.4.1 (Geometric moment generating function). Let X be a geometric random variable with parameter p. The moment generating function for X is given by

$$m_X(t) = \frac{pe^t}{1 - qe^t} \qquad t < -\ln q$$

where $q = 1 - p$.

Proof. The density for X is given by

$$f(x) = q^{x-1}p \qquad x = 1, 2, 3, \ldots$$

By definition

$$\begin{aligned} m_X &= E[e^{tX}] \\ &= \sum_{\text{all } x} e^{tx} f(x) \\ &= \sum_{x=1}^{\infty} e^{tx} q^{x-1} p \\ &= pq^{-1} \sum_{x=1}^{\infty} (qe^t)^x \end{aligned}$$

The series on the right is a geometric series with first term qe^t and common ratio qe^t. Thus

$$\begin{aligned} m_X(t) &= pq^{-1}\left(\frac{qe^t}{1 - qe^t}\right) \\ &= \frac{pe^t}{1 - qe^t} \end{aligned}$$

provided $|r| = |qe^t| < 1$. Since the exponential function is nonnegative and $0 < q < 1$, this restriction implies that $qe^t < 1$. The inequality is solved for t as follows:

$$\begin{aligned} qe^t &< 1 \\ e^t &< 1/q \\ \ln e^t &< \ln 1/q \\ t &< \ln 1 - \ln q \\ t &< -\ln q \end{aligned}$$

The following theorem shows how the moment generating function can be used to generate ordinary moments for a random variable X.

Theorem 3.4.2. Let $m_X(t)$ be the moment generating function for a random variable X. Then

$$\left.\frac{d^k m_X(t)}{dt^k}\right|_{t=0} = E[X^k]$$

Proof. Recall from elementary calculus that the Maclaurin series expansion for e^z is

$$e^z = 1 + z + z^2/2! + z^3/3! + z^4/4! + \cdots$$

Letting $z = tX$, the Maclaurin series expansion for e^{tX} is

$$e^{tX} = 1 + tX + (tX)^2/2! + (tX)^3/3! + (tX)^4/4! + \cdots$$

By taking the expected value of each side of this equation, we obtain

$$\begin{aligned} m_X(t) = E[e^{tX}] &= E[1 + tX + t^2X^2/2! + t^3X^3/3! + t^4X^4/4! + \cdots] \\ &= 1 + tE[X] + t^2/2!E[X^2] + t^3/3!E[X^3] + t^4/4!E[X^4] + \cdots \end{aligned}$$

Differentiating this series term by term with respect to t, we see that

$$\frac{dm_X(t)}{dt} = E[X] + tE[X^2] + t^2/2!E[X^3] + t^3/3!E[X^4] + \cdots$$

When this derivative is evaluated at $t = 0$, every term except the first becomes 0. Hence

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = E[X]$$

Taking the second derivative of $m_X(t)$, we obtain

$$\frac{d^2m_X(t)}{dt^2} = E[X^2] + tE[X^3] + t^2/2!E[X^4] + \cdots$$

Evaluating this derivative at $t = 0$ yields

$$\left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = E[X^2].$$

This procedure can be continued to show that

$$\left.\frac{d^km_X(t)}{dt^k}\right|_{t=0} = E[X^k]$$

for any positive integer k as desired.

Let us use the moment generating function to find a general expression for the mean and variance of a geometric distribution with parameter p.

Theorem 3.4.3. Let X be a geometric random variable with parameter p. Then

$$E[X] = 1/p \qquad \text{and} \qquad \operatorname{Var} X = q/p^2$$

Proof. For a geometric random variable with parameter p

$$m_X(t) = \frac{pe^t}{1 - qe^t}$$

$$\frac{dm_X(t)}{dt} = \frac{(1 - qe^t)pe^t + pe^tqe^t}{(1 - qe^t)^2}$$

$$= \frac{pe^t}{(1 - qe^t)^2}$$

Evaluating this derivative at $t = 0$, we obtain

$$E[X] = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{p}{(1-q)^2}$$
$$= p/p^2$$
$$= 1/p$$

Taking the second derivative of $m_X(t)$, we obtain

$$\frac{d^2 m_X(t)}{dt^2} = \frac{(1-qe^t)^2 pe^t + 2pe^t(1-qe^t)qe^t}{(1-qe^t)^4}$$
$$= \frac{pe^t(1-qe^t)\left[(1-qe^t) + 2qe^t\right]}{(1-qe^t)^4}$$
$$= \frac{pe^t(1+qe^t)}{(1-qe^t)^3}$$

Evaluating this derivative at $t = 0$, we see that

$$E[X^2] = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{p(1+q)}{(1-q)^3} = \frac{(1+q)}{p^2}$$

Now

$$\operatorname{Var} X = E[X^2] - (E[X])^2$$
$$= \frac{1+q}{p^2} - \frac{1}{p^2}$$
$$= \frac{q}{p^2}$$

We illustrate the use of these theorems by finding the moment generating function, mean and variance for the random variable of Example 3.4.1.

Example 3.4.3. Consider the random variable X, the number of trials needed to obtain the first zero when generating a series of random digits. Since this random variable is geometric with parameter $p = 1/10$

$$m_X(t) = \frac{pe^t}{1-qe^t} = \frac{(1/10)e^t}{1-(9/10)e^t}$$
$$\mu = E[X] = 1/p = 10$$
$$\sigma^2 = \operatorname{Var} X = q/p^2 = \frac{9/10}{(1/10)^2} = 90$$

Note that this value for μ agrees with that obtained in Example 3.4.2.

The importance of the moment generating function for a random variable is not completely evident at this time. It does give us a way to find general expressions for the mean and variance as well as for the ordinary moments of an entire family of random variables. As we shall see later, the moment generating function, when it exists, serves as a fingerprint that completely identifies the random variable under study. This function will be used extensively in the remainder of this text.

3.5 BINOMIAL DISTRIBUTION

The next distribution to be studied is the *binomial* distribution. Once again, you have already seen some binomial random variables even though they were not labeled as such at the time. The theoretical basis for working with this distribution is the binomial theorem presented in most beginning algebra courses. The statement of this theorem is as follows:

Binomial theorem. For any two real numbers a and b and any positive integer n

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

where $\binom{n}{k}$ is given by $\dfrac{n!}{k!(n-k)!}$

To recognize a situation that involves a binomial random variable, you must be familiar with the assumptions that underlie this distribution. These are given below:

Binomial properties

1. The experiment consists of a *fixed* number, n, of Bernoulli trials, trials that result in either a "success" (s) or a "failure" (f).
2. The trials are identical and independent and therefore the probability of success, p, remains the same from trial to trial.
3. The random variable X denotes the number of successes obtained in the n trials.

Once we realize that the binomial model is appropriate from the physical description of the experiment, we will want to describe the behavior of the binomial random variable involved. To do so, we need to consider the density for the random variable. To get an idea of the general form for the binomial density, let us consider the case in which $n = 3$. The sample space for such an experiment is

$$S = \{fff, sff, fsf, ffs, ssf, sfs, fss, sss\}$$

Since the trials are independent, the probability assigned to each sample point is found by multiplying. For example, the probabilities assigned to the sample points fff and sff are $(1-p)(1-p)(1-p) = (1-p)^3$ and $p(1-p)(1-p) = p(1-p)^2$, respectively. The random variable X assumes the value 0 only if the experiment results in the outcome fff. That is,

$$P[X = 0] = (1-p)^3$$

However, X assumes the value 1 if the experiment results in any one of the outcomes *sff*, *fsf*, or *ffs*. Thus

$$P[X = 1] = 3 \cdot p(1-p)^2$$

Similarly,

$$P[X = 2] = 3 \cdot p^2(1-p)$$

and

$$P[X = 3] = p^3$$

It is evident that for $x = 0, 1, 2, 3$

$$P[X = x] = c(x)p^x(1-p)^{3-x}$$

where $c(x)$ denotes the number of sample points that correspond to x successes. Such a sample point is expressed as a permutation of three letters with x of these being s's and the rest, $3 - x$, of these being f's. Using the formula for the number of permutations of indistinguishable objects studied in Chap. 1, we see that

$$c(x) = \frac{3!}{x!(3-x)!} = \binom{3}{x}$$

Thus the density for this binomial random variable is given by

$$f(x) = \binom{3}{x} p^x(1-p)^{3-x} \qquad x = 0, 1, 2, 3$$

To generalize this idea to n trials, we replace 3 by n to obtain the expression

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

This suggests the formal definition of the binomial distribution.

Definition 3.5.1 (Binomial distribution). A random variable X has a binomial distribution with parameters n and p if its density is given by

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$
$$0 < p < 1$$

where n is a positive integer.

To see that the function given in this definition is a density, note that it is nonnegative. Furthermore, by applying the binomial theorem with $k = x$, $a = p$, and $b = 1 - p$ it can be seen that

$$\sum_{x=0}^{n} \binom{n}{x} p^x(1-p)^{n-x} = [p + (1-p)]^n = 1$$

as desired.

Example 3.5.1. Recent studies of German air traffic controllers have shown that it is difficult to maintain accuracy when working for long periods of time on data display screens. A surprising aspect of the study is that the ability to detect spots on a radar screen decreases as their appearance becomes too rare. The probability of correctly identifying a signal is approximately .9 when 100 signals arrive per 30-minute period. This probability drops to .5 when only 10 signals arrive at random over a 30-minute period. The hypothesis is that unstimulated minds tend to wander. Let X denote the number of signals correctly identified in a 30-minute time span in which 10 signals arrive. This experiment consists of a series of $n = 10$ independent and identical Bernoulli trials with "success" being the correct identification of a signal. The probability of success is $p = 1/2$. Since X denotes the number of successes in a fixed number of trials, X is binomial. Its density is found by letting $n = 10$ and $p = 1/2$ in the expression for f given in Definition 3.5.1. That is,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

or

$$f(x) = \binom{10}{x} (1/2)^x (1/2)^{10-x} \qquad x = 0, 1, 2, \ldots, 10$$

(Based on a study reported in "Human Aspects of Quality Assurances" by W. E. Masing, *Quality Assurance*, volume 8, no. 2, June, 1982, p. 35.)

The next theorem summarizes other theoretical properties of the binomial distribution. Its proof is left as an exercise (Exercise 41).

Theorem 3.5.1. Let X be a binomial random variable with parameters n and p.

1. The moment generating function for X is given by

$$m_X(t) = (q + pe^t)^n \qquad q = 1 - p$$

2. $E[X] = \mu = np$
3. $\operatorname{Var} X = \sigma^2 = npq$

Example 3.5.2. The random variable X, the number of radar signals properly identified in a 30-minute period, is a binomial random variable with parameters $n = 10$ and $p = 1/2$. The moment generating function for this random variable is

$$m_X(t) = (1/2 + 1/2e^t)^{10}$$

Its mean is $\mu = np = 10(1/2) = 5$ and its variance is $\sigma^2 = npq = 10(1/2)(1/2) = 10/4$.

In statistical studies we shall usually be interested in computing the probability that the random variable assumes certain values. This probability can be computed from the density function, f, or from the cumulative distribution function, F. Since the binomial distribution comes into play in such a wide variety of physical applications, tables of the cumulative distribution function for selected values of n and p have been compiled. Table I of App. A is one such table. That is, Table I gives the values of

$$F(t) = \sum_{x=0}^{[t]} \binom{n}{x} p^x (1-p)^{n-x}$$

for selected values of n and p where $[t]$ represents the greatest integer less than or equal to t. Its use is illustrated in the following example.

Example 3.5.3. Let X denote the number of radar signals properly identified in a 30-minute time period in which 10 signals are received. Assuming that X is binomial with $n = 10$ and $p = 1/2$, find the probability that at most seven signals will be identified correctly. This probability can be found by summing the density from $x = 0$ to $x = 7$. That is,

$$P[X \leq 7] = \sum_{x=0}^{7} \binom{10}{x} (1/2)^x (1/2)^{10-x}$$

Evaluating this probability directly entails a large amount of arithmetic. However, its value can be read from Table I of App. A. We first look at the group of values labeled $n = 10$. The desired probability of .9453 is found in the column labeled .5 and the row labeled 7. That is,

$$P[X \leq 7] = F(7) = .9453$$

Other probabilities can be found by first rewriting the question in terms of the cumulative distribution function. For example,

$$\begin{aligned} P[2 \leq X \leq 7] &= P[X \leq 7] - P[X < 2] \\ &= P[X \leq 7] - P[X \leq 1] \\ &= F(7) - F(1) \\ &= .9453 - .0107 \\ &= .9346 \end{aligned}$$

Later in the text we shall show ways of approximating binomial probabilities when the values of n and p are such that no appropriate binomial table is available.

3.6 NEGATIVE BINOMIAL DISTRIBUTION (OPTIONAL)

The negative binomial distribution is a distribution that can be thought of as a "reversal" of the binomial distribution. In the binomial setting, the random variable X represents the number of successes obtained in a series of n independent and identical Bernoulli trials; the number of trials is *fixed* and the number of successes will *vary* from experiment to experiment. The negative binomial random variable represents the number of trials needed to obtain exactly r successes; here, the number of successes is *fixed* and the number of trials will *vary* from experiment to experiment. In particular, the negative binomial random variable arises in situations characterized by these properties:

Negative binomial properties

1. The experiment consists of a series of independent and identical Bernoulli trials each with probability p of success.
2. The trials are observed until exactly r successes are obtained where r is fixed by the experimenter.
3. The random variable X is the number of trials needed to obtain the r successes.

It is not hard to derive the density function for X. To do so, let us consider a setting in which $r = 3$. Typical outcomes for such an experiment are

$$ssffffs \quad sffffss \quad ffffsss \quad sss \quad ssfs$$

Here X assumes the values 7, 7, 7, 3, and 4, respectively. There are several things to notice immediately. First, each outcome must end with a successful trial. Second, the remaining $x - 1$ trials must result in exactly 2 successes and $x - 3$ failures *in some order*. Third, different outcomes can yield identical values for X. To determine the number of outcomes that result in a given value of X, we ask: "How many permutations can be formed consisting of $x - 1$ objects of which exactly 2 represent success and the rest, $x - 3$, represent failure?" Exercise 17 of Chap. 1 can be applied to see that the answer to this question is $\binom{x-1}{2}$. For example, there are $\binom{6}{2} = 15$ ways in which X can assume the value 7. Three of these outcomes are given above. Since trials are independent with probability p of success and probability $1 - p$ of failure, the probability of an outcome for which $X = x$ is given by

$$P[X = x] = \binom{x-1}{2}(1-p)^{x-3}p^3 \qquad x = 3, 4, 5, \ldots$$

You can use this expression to verify that the probability that $X = 7$ is

$$\binom{6}{2}(1-p)^4 p^3$$

The argument given for $r = 3$ can be generalized easily. We simply replace 3 by r and 2 by $r - 1$ in the argument given to obtain the following definition for the negative binomial random variable.

Definition 3.6.1 (Negative binomial distribution). A random variable X is said to have a negative binomial distribution with parameters p and r if its density f is given by

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r \qquad r = 1, 2, 3, \ldots$$
$$x = r, r+1, r+2, \ldots$$

A Taylor series expansion in q where $q = 1 - p$ is used to show that this density sums to 1 as required. The proof is a bit tricky and is outlined in Exercise 45.

Theorem 3.6.1 gives the moment generating function for the negative binomial distribution. Its derivation is also based on a Taylor series expansion. (See Exercise 46.) The expectations stated in the theorem are obtained from the moment generating function.

Theorem 3.6.1. Let X be a negative binomial random variable with parameters r and p. Then

1. the moment generating function for X is given by

$$m_X(t) = \frac{(pe^t)^r}{(1 - qe^t)^r} \qquad q = 1 - p$$

2. $E[X] = r/p$
3. $\mathrm{Var}(X) = rq/p^2$

An example will illustrate the use of this distribution in a practical setting.

Example 3.6.1. Cotton linters used in the production of rocket propellant are subjected to a nitration process that enables the cotton fibers to go into solution. The process is 90% effective in that the material produced can be shaped as desired in a later processing stage with probability .9. What is the probability that exactly 20 lots will be produced in order to obtain the third defective lot? Here "success" is obtaining a defective lot and hence $p = .1$ and $r = 3$. The probability that $X = 20$ is given by

$$f(20) = \binom{19}{2}(.9)^{17}(.1)^3$$

The expected value of X is $r/p = 3/.1$ or 30 and the variance of X is $rq/p^2 = 3(.9)/(.1)^2 = 270$. (Based on a study to compare different sources of cotton linters conducted by the Radford University Statistical Consulting Service for the Radford Army Ammunition Plant.)

One other point should be made. When $r = 1$, the negative binomial distribution reduces to the geometric distribution studied earlier. (See Exercise 49.)

3.7 HYPERGEOMETRIC DISTRIBUTION

Sampling from a finite population can be done in one of two ways. An item can be selected, examined, and returned to the population for possible reselection; or it can be selected, examined, and kept, thus preventing its reselection in subsequent draws. The former is called *sampling with replacement* whereas the latter *is sampling without replacement*. Sampling with replacement guarantees that the draws are independent. In sampling without replacement, the draws are *not* independent. Thus, if we sample without replacement, the random variable X, the number of successes in n draws, is no longer binomial. Rather, it follows a distribution known as the *hypergeometric distribution*.

To derive the density for this distribution, suppose that we have a group of N objects and that r of these objects have a trait of interest to us. We are to select

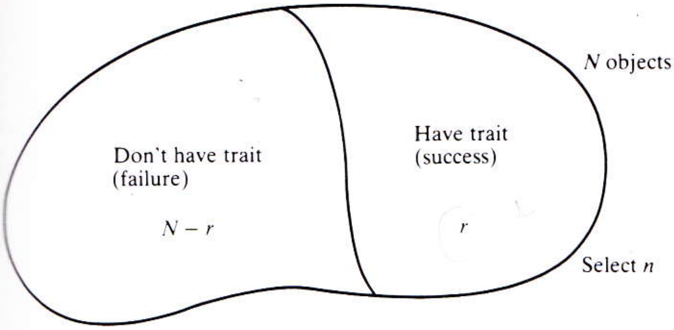


FIGURE 3.1

n objects from the group randomly without replacement. Let X denote the number of objects chosen that have the trait. The idea is depicted in Fig. 3.1. Since we are not interested in the order in which items are selected, we can use combinatorial techniques to conclude that there are $\binom{N}{n}$ ways to choose the n objects. In a random selection, we are just as likely to obtain one set of n objects as any other. That is, there are $\binom{N}{n}$ equally likely ways in which this experiment can proceed. In order to have x successes, we must select exactly x objects from the r objects with the trait of interest; this can be done in $\binom{r}{x}$ ways. We must select the remaining $n - x$ objects from the $N - r$ objects that do not have the trait; this can be done in $\binom{N-r}{n-x}$ ways. Using classical probability and the multiplication rule for counting

$$P[X = x] = \frac{\text{number of ways to select } x \text{ objects with the trait and } n - x \text{ objects without the trait}}{\text{number of ways the experiment can proceed}}$$

$$= \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

This argument suggests the definition of the hypergeometric distribution.

Definition 3.7.1 (Hypergeometric distribution). A random variable X has a hypergeometric distribution with parameters N, n, and r if its density is given by

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \qquad \max[0, n - (N - r)] \le x \le \min(n, r)$$

where N, r, and n are positive integers.

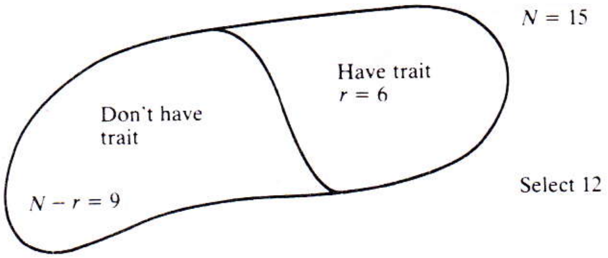


FIGURE 3.2

Notice the unusual bounds for X. A simple numerical example should show you why these bounds are as stated.

Example 3.7.1. Suppose that X is hypergeometric with $N = 15$, $r = 6$, and $n = 12$. This situation is depicted in Fig. 3.2. Since only six items have the desired trait, X cannot exceed 6. Note that $6 = \min(n, r) = \min(12, 6)$. Since we can select at most nine items from among those without the trait, we must select at least three items from among those with the trait. Note that

$$3 = \max[0, n - (N - r)] = \max[0, 12 - (15 - 6)] = \max[0, 3]$$

Just be careful when stating the bounds for a hypergeometric random variable. They are tricky! Since the bounds for X are unusual, the theoretical development of the hypergeometric distribution is not easy. However, it can be shown that

$$E[X] = n\left(\frac{r}{N}\right) \qquad \text{and} \qquad \operatorname{Var} X = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Example 3.7.2. A foundry ships engine blocks in lots of size 20. Since no manufacturing process is perfect, defective blocks are inevitable. However, to detect the defect, the block must be destroyed. Thus, we cannot test each block. Before accepting a lot, three items are selected and tested. Suppose that a given lot actually contains five defective items. Let X denote the number of defective items sampled. The density for X is

$$f(x) = \frac{\binom{5}{x}\binom{15}{3-x}}{\binom{20}{3}} \qquad x = 0, 1, 2, 3$$

The expected number of defective blocks in a sample of size 3 is

$$E[X] = n\left(\frac{r}{N}\right) = 3\left(\frac{5}{20}\right) = \frac{3}{4}$$

The variance for X is

$$\begin{aligned}\operatorname{Var} X &= n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)\\ &= 3\left(\frac{5}{20}\right)\left(\frac{15}{20}\right)\left(\frac{17}{19}\right)\\ &= \frac{153}{304}\end{aligned}$$

If the number of items sampled (n) is small relative to the number of objects from which the sample is drawn (N), then the binomial distribution can be used to approximate hypergeometric probabilities. A rule of thumb is that the approximation is usually satisfactory if $n/N \leq .05$. The proof of this result depends upon Stirling's formula which is studied in courses in advanced calculus. We shall not attempt the proof here. However, the result should not be surprising. If n is small relative to N, then the composition of the sampled group does not change much from trial to trial even though we are keeping the sampled items. Thus, the probability of success is not changing much from trial to trial and for all practical purposes it can be viewed as being constant. Thus the distribution of X, the number of successes obtained in n draws, can be approximated by the binomial distribution with parameters n and $p = r/N$.

Example 3.7.3. During the course of an hour, one thousand bottles of beer are filled by a particular machine. Each hour a sample of 20 bottles is randomly selected and the number of ounces of beer per bottle is checked. Let X denote the number of bottles selected that are underfilled. Suppose that, during a particular hour, 100 underfilled bottles are produced. Find the probability that at least three underfilled bottles will be among those sampled. The exact value of this probability is given by

$$\begin{aligned}P[X \geq 3] &= 1 - P[X < 3]\\ &= 1 - P[X \leq 2]\\ &= 1 - P[X = 0] - P[X = 1] - P[X = 2]\\ &= 1 - \frac{\binom{100}{0}\binom{900}{20}}{\binom{1000}{20}} - \frac{\binom{100}{1}\binom{900}{19}}{\binom{1000}{20}} - \frac{\binom{100}{2}\binom{900}{18}}{\binom{1000}{20}} = .3224\end{aligned}$$

As you can see, calculating this probability directly, even with the aid of a calculator, is time-consuming. However, since $n/N = 20/1000 \leq .05$ our rule of thumb indicates that this probability can be approximated using the binomial distribution with parameters $n = 20$ and $p = r/N = 100/1000 = .1$. From Table I of the Appendix, the cumulative binomial table, we find that

$$\begin{aligned}P[X \geq 3] &= 1 - P[X < 3]\\ &= 1 - P[X \leq 2]\\ &= 1 - .6769\\ &= .3231\end{aligned}$$

3.8 POISSON DISTRIBUTION

The last discrete family to be considered is the family of *Poisson* random variables, named for the French mathematician Simeon Denis Poisson (1781–1840). The Maclaurin series expansion for the function e^z studied in beginning calculus courses provides the theoretical basis for this distribution. This series is given by

Maclaurin series. For z a real number

$$e^z = 1 + z + z^2/2! + z^3/3! + z^4/4! + \cdots$$

We begin by considering the mathematical properties of this important family of random variables.

Definition 3.8.1 (Poisson distribution). A random variable X is said to have a Poisson distribution with parameter k if its density f is given by

$$f(x) = \frac{e^{-k}k^x}{x!} \qquad \begin{array}{l} x = 0, 1, 2, \ldots \\ k > 0 \end{array}$$

The function f given in this definition is nonnegative. To see that it sums to one, note that

$$\sum_{x=0}^{\infty} \frac{e^{-k}k^x}{x!} = e^{-k}(1 + k + k^2/2! + k^3/3! + \cdots)$$

The series on the right is the Maclaurin series for e^k. Thus

$$\sum_{x=0}^{\infty} \frac{e^{-k}k^x}{x!} = e^{-k}e^k = e^0 = 1$$

as desired.

The moment generating function for this distribution is easy to obtain as is its mean and variance. The following theorem gives these results. Its proof is outlined as an exercise. (Exercise 67.)

Theorem 3.8.1. Let X be a Poisson random variable with parameter k.

1. The moment generating function for X is given by

$$m_X(t) = e^{k(e^t - 1)}$$

2. $E[X] = k$
3. $\operatorname{Var} X = k$

Poisson random variables usually arise in connection with what are called *Poisson processes*. Poisson processes involve observing discrete events in a continuous "interval" of time, length, or space. We use the word "interval" in describing the general Poisson process with the understanding that we may not be dealing with an interval in the usual mathematical sense. For example, we might observe the number of white blood cells in a drop of blood. The discrete event of interest is the observation of a white cell, whereas the continuous "interval" involved is a drop of blood. We might observe the number of times radioactive gases are emitted from a nuclear power plant during a three-month period. The discrete event of concern is the emission of radioactive gases. The continuous interval consists of a period of three months. The variable of interest in a Poisson process is X, the number of occurrences of the event in an interval of length s units. Although the derivation is a bit tricky, it can be shown using differential equations that X is a Poisson random variable with parameter $k = \lambda s$ where λ is a positive number that characterizes the underlying Poisson process. To understand the physical significance of the constant λ, note that by Definition 3.8.1 the density for X is given by

$$f(x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!} \qquad x = 0, 1, 2, 3, \ldots$$

By Theorem 3.8.1, the expected value of X is λs. That is, the average number of occurrences of the event of interest in an interval of s units is λs. Thus, the average number of occurrences of the event in one unit of time, length, area, or space is $\lambda s/s = \lambda$. That is, physically, the *parameter* λ *of a Poisson process represents the average number of occurrences of the event in question per measurement unit.*

These concepts are illustrated in Example 3.8.1.

Example 3.8.1. The white blood-cell count of a healthy individual can average as low as 6000 per cubic millimeter of blood. To detect a white-cell deficiency, a .001 cubic millimeter drop of blood is taken and the number of white cells X is found. How many white cells are expected in a healthy individual? If at most two are found, is there evidence of a white-cell deficiency?

This experiment can be viewed as involving a Poisson process. The discrete event of interest is the occurrence of a white cell; the continuous interval is a drop of blood.

Let the measurement unit be a cubic millimeter; then $s = .001$ and λ, the average number of occurrences of the event per unit, is 6000. Thus X is a Poisson random variable with parameter $\lambda s = 6000(.001) = 6$. By Theorem 3.8.1, $E[X] = \lambda s = 6$. In a healthy individual, we would expect, on the average, to see six white cells. How rare is it to see at most two? That is, what is $P[X \le 2]$? From Definition 3.8.1

$$P[X \le 2] = \sum_{x=0}^{2} f(x) = \sum_{x=0}^{2} \frac{e^{-6}6^x}{x!}$$

$$= \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!}$$

Evaluating this type of expression directly does entail some arithmetic.

Once again, because of the wide appeal of the Poisson model, the values of the cumulative distribution function for selected values of the parameter $k = \lambda s$ are tabulated. Table II of App. A is one such table. The desired probability of .062 is found by looking under the column labeled $k = 6$ in the row labeled 2. Is there evidence of a white-cell deficiency? There are no rules that say at what point probabilities are considered to be small. To answer this question, a value judgment must be made. If you consider .062 to be small, then the natural conclusion is that the individual does have a white-cell deficiency.

One other important application of the Poisson density should be mentioned. Occasionally, one encounters a binomial random variable for which n is very large. If binomial tables do not exist for this value of n, then calculating probabilities associated with the variable becomes cumbersome. It is often possible to approximate these binomial probabilities via an appropriately chosen Poisson distribution. The following theorem, first presented in 1837 by Poisson, shows that as n becomes large and p becomes small, the binomial density approaches that of the Poisson. Its proof is based upon the following result usually proved in courses in advanced calculus [30]

$$\lim_{n \to \infty} \left(1 + \frac{z}{n}\right)^n = e^z \qquad \text{for } z \text{ real}$$

Theorem 3.8.2 (Poisson approximation to the binomial distribution). Let X be binomial with parameters n and p. If the parameter n approaches infinity and p approaches 0 in such a way that np remains constant at some value $k > 0$, then

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \doteq \frac{e^{-k} k^x}{x!}$$

Proof. Let X be binomial with parameters n and p and assume that $np = k > 0$. By Definition 3.5.1

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$$

Let $k = np$ or $p = k/n$, then

$$\begin{aligned}
P[X = x] &= \binom{n}{x} (k/n)^x (1 - k/n)^{n-x} \\
&= \frac{n!}{x!(n-x)!} \frac{k^x}{n^x} (1 - k/n)^n \frac{1}{(1 - k/n)^x} \\
&= \frac{k^x}{x!} \frac{n(n-1)(n-2)\cdots(n-x+1)(n-x)!}{n \cdot n \cdot n \cdots n(n-x)!} \frac{(1-k/n)^n}{(1-k/n)^x} \\
&= k^x/x!(1)(1 - 1/n)(1 - 2/n)\cdots\left(1 - \frac{x-1}{n}\right)\frac{(1-k/n)^n}{(1-k/n)^x}
\end{aligned}$$

Now let n approach infinity to obtain

$$P[X = x] \doteq \lim_{n\to\infty} k^x/x! \frac{(1)(1 - 1/n)(1 - 2/n)\cdots\left(1 - \frac{x-1}{n}\right)}{(1 - k/n)^x}(1 - k/n)^n$$

$$= k^x/x! \lim_{n\to\infty} \frac{(1)(1 - 1/n)(1 - 2/n)\cdots\left(1 - \frac{x-1}{n}\right)}{(1 - k/n)^x}$$

$$\times \lim_{n\to\infty} [1 + (-k/n)]^n$$

$$= \frac{e^{-k}k^x}{x!}$$

Note that this theorem states that the desired binomial probability is approximately equal to a Poisson probability with the Poisson parameter $k = n \cdot p$. This approximation is usually good if $n \geq 20$ and $p \leq .05$ and very good if $n \geq 100$ and $np \leq 10$. Since the approximation is used when n is large and p is small, the Poisson distribution is often called the distribution of "rare" events. Its use is illustrated in the next example.

Example 3.8.2. In manufacturing electronic circuits, ceramic plates are drilled to provide pathways called "vias" from one surface to another. A typical plate is about the size of a playing card and may require 10,000 vias each as small as a pinpoint. In the past these vias were drilled using diamond drills. New technology uses lasers to produce these precisely positioned pathways. Suppose that the probability of incorrectly positioning a via is only 1/20,000. What is the probability that a randomly selected plate will have no improperly positioned vias?

Let X denote the number of vias that are positioned incorrectly. X is binomial with $n = 10{,}000$ and $p = 1/20{,}000$. We want to approximate

$$P[X = 0] = \binom{10{,}000}{0}(1/20{,}000)^0(19{,}999/20{,}000)^{10{,}000} = .6065$$

Since n is large and p is small, this binomial probability can be approximated by a Poisson probability with the Poisson parameter $k = np = 10{,}000(1/20{,}000) = .5$. That is,

$$P[X = 0] \doteq \frac{e^{-.5}(.5)^0}{0!}$$

From Table II of App. A, this probability is approximately .607. (Based on a report in *Laser Focus*, December, 1982, p. 26.)

Table 3.6 summarizes the discrete distributions that have been presented.

TABLE 3.6

Discrete distributions: a summary

Name	Density		Moment generating function	Mean	Variance
Geometric	$(1-p)^{x-1}p$	$x = 1, 2, 3, \ldots$ $0 < p < 1$	$\dfrac{pe^t}{1-qe^t}$	$1/p$	q/p^2
Uniform	$1/n$	$x = x_1, x_2, \ldots, x_n$ n a positive integer	$\dfrac{\sum_{i=1}^n e^{tx_i}}{n}$	$\dfrac{\sum_{i=1}^n x_i}{n}$	$\dfrac{\sum_{i=1}^n x_i^2}{n} - \left(\dfrac{\sum_{i=1}^n x_i}{n}\right)^2$
Binomial	$\binom{n}{x} p^x(1-p)^{n-x}$	$x = 0, 1, 2, \ldots, n$ $0 < p < 1$ n a positive integer	$(q + pe^t)^n$	np	$np(1-p)$
Bernoulli	$p^x(1-p)^{1-x}$	$x = 0, 1$ $0 < p < 1$	$q + pe^t$	p	$p(1-p)$
Hypergeometric	$\dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\max[0, n-(N-r)]$ $\le x \le \min(n, r)$		$n\dfrac{r}{N}$	$n\dfrac{r}{N}\left(\dfrac{N-r}{N}\right)\left(\dfrac{N-n}{N-1}\right)$
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$x = r, r+1, r+2, \ldots,$ $0 < p < 1$	$\dfrac{(pe^t)^r}{(1-qe^t)^r}$	r/p	$\dfrac{r(1-p)}{p^2}$
Poisson	$\dfrac{e^{-k}k^x}{x!}$	$x = 0, 1, 2, \ldots$ $k > 0$	$e^{k(e^t-1)}$	k	k

3.9 SIMULATING A DISCRETE DISTRIBUTION (OPTIONAL)

In designing operating systems of various types, it is often necessary to simulate the system before it is built. Simulation is usually done with the aid of a computer. However, the idea behind simulation can be illustrated using a random digit table. A portion of such a table is given in Table III of App. A. Its use is illustrated in the following example.

Example 3.9.1. Table 3.7 presents a portion of the random digit table in the appendix. Let us read a sequence of random two-digit numbers from this table. To do so we must get a random start. This can be done by writing the integers 1 through 14 on slips of paper, placing the slips in a bowl, stirring, and drawing one slip at random from the bowl. The number selected identifies the column in which our starting number is located. In a similar way we can select the row in which the starting number is located. Suppose that this process results in the selection of column 2 and row 5. This identifies the random starting point as 39975.

Since we want two-digit numbers, we need only read the first two digits of this number. Thus our first random number is 39. Since a random digit table is constructed in such a way that the digit appearing at each position in the table is just as likely to be one digit as any other, the table can be read in any way. Let us agree to read down the second column so that the next four two-digit numbers are 06, 72, 91, and 14.

The next example illustrates the use of a random digit table in a simple simulation experiment.

Example 3.9.2. Suppose that at a particular airport planes arrive at an average rate of one per minute and depart at the same average rate. We are interested in simulating the behavior of the random variable Z, the number of planes on the ground at a given time. We will simulate Z for five consecutive one-minute periods.

TABLE 3.7

Row \ Column	Random digits (1)	(2)	(3)
1	10480	15011	01536
2	22368	46573	25595
3	24130	48360	22527
4	42167	93093	06243
5	37570	39975	81837
6	77921	06907	11008
7	99562	72905	56420
8	96301	91977	05463
9	89579	14342	63661
10	85485	36857	43342

TABLE 3.8

Random number	Number of arrivals (x)	Number of departures (y)	$P[X = x] = P[Y = y]$
000–367	0	0	.368
368–735	1	1	.368
736–919	2	2	.184
920–980	3	3	.061
981–995	4	4	.015
996–998	5	5	.003
999	6	6	.001

Note that for each of these periods the random variables X, the number of arrivals, and Y, the number of departures, are each Poisson variables with parameter $k = 1$. The density for X and Y is obtained from Table II of App. A and is shown below.

$$P[X = 0] = P[Y = 0] = .368$$
$$P[X = 1] = P[Y = 1] = .368$$
$$P[X = 2] = P[Y = 2] = .184$$
$$P[X = 3] = P[Y = 3] = .061$$
$$P[X = 4] = P[Y = 4] = .015$$
$$P[X = 5] = P[Y = 5] = .003$$
$$P[X = 6] = P[Y = 6] = .001$$
$$P[X > 6] = P[Y > 6] \doteq 0$$

There are 1000 possible three-digit numbers. We divide them into seven categories to reflect the above probabilities. This division is shown in Table 3.8. To perform the simulation we read a total of 10 random three-digit numbers using the procedure demonstrated in Example 3.9.1. Assume that at the beginning of the simulation

TABLE 3.9

Time span, min	Random 3-digit number	Number of arrivals (x)	Number of departures (y)	Number on ground at end of time period (z)
1	015	0		100
	255		0	100
2	225	0		
	062		0	100
3	818	2		
	110		0	102
4	564	1		
	054		0	103
5	636	1		
	433		1	103

there are 100 planes on the ground and that our random starting point is the number 01536 found in line 1 and column 3 of Table 3.7. The first number read corresponds to the arrivals during the first minute of observation, the second to the departures during this time span, and so forth. The results of the simulation are shown in Table 3.9. If this simulation were continued over a long period of time, we could begin to answer such questions as: "On the average how many planes are on the ground at a given time?", and "How much variability is there in the number of planes on the ground?"

CHAPTER SUMMARY

In this chapter we introduced the concept of a random variable and showed you how to distinguish a discrete random variable from one that is not discrete. We studied two functions, the density function and the cumulative distribution function, that are used to compute probabilities. The density gives the probability that X assumes a specific value x; the cumulative distribution gives the probability that X assumes a value less than or equal to x. The concept of expected value was introduced and used to define three important parameters, the mean (μ), the variance (σ^2), and the standard deviation (σ). The mean is a measure of the center of location of the distribution; the variance and standard deviation measure the variability of the random variable about its mean. The moment generating function was introduced as a means of finding the mean and variance of X. Special discrete distributions that find extensive use in all areas of application were presented. These are the geometric, hypergeometric, negative binomial, binomial, Bernoulli, uniform, and Poisson distributions. We also discussed briefly how to simulate a discrete distribution. These terms were defined:

Random variable	Variance
Discrete random variable	Standard deviation
Discrete density	Bernoulli trial
Cumulative distribution	Moment generating function
Expected value	Sampling with replacement
Mean	Sampling without replacement

EXERCISES

Section 3.1

In each of the following, identify the variable as discrete or not discrete.

1. T: the turnaround time for a computer job (the time it takes to run the program and receive the results).
2. M: the number of meteorites hitting a satellite per day.
3. N: the number of neutrons expelled per thermal neutron absorbed in fission of uranium-235.
4. Neutrons emitted as a result of fission are either prompt neutrons or delayed neutrons. Prompt neutrons account for about 99% of all neutrons emitted and are

released within 10^{-14} s of the instant of fission. Delayed neutrons are emitted over a period of several hours. Let D denote the time at which a delayed neutron is emitted in a fission reaction.

5. Electrical resistance is the opposition which is offered by electrical conductors to the flow of current. The unit of resistance is the ohm. For example, a $2\frac{1}{2}$-inch electric bell will usually have a resistance somewhere between 1.5 and 3 ohms. Let O denote the actual resistance of a randomly selected bell of this type.

6. The number of power failures per month in the Tennessee Valley power network.

Section 3.2

7. Grafting, the uniting of the stem of one plant with the stem or root of another is widely used commercially to grow the stem of one variety that produces fine fruit on the root system of another variety with a hardy root system. Most Florida sweet oranges grow on trees grafted to the root of a sour orange variety. The density for X, the number of grafts that fail in a series of five trials, is given by Table 3.10.

TABLE 3.10

x	0	1	2	3	4	5
$f(x)$	.7	.2	.05	.03	.01	? 0 21

(*a*) Find $f(5)$.
(*b*) Find the table for F.
(*c*) Use F to find the probability that at most three grafts fail; that at least two grafts fail.
(*d*) Use F to verify that the probability of exactly three failures is .03.

8. In blasting soft rock such as limestone, the holes bored to hold the explosives are drilled with a Kelly bar. This drill is designed so that the explosives can be packed into the hole before the drill is removed. This is necessary since in soft rock the hole often collapses as the drill is removed. The bits for these drills must be changed fairly often. Let X denote the number of holes that can be drilled per bit. The density for X is given in Table 3.11. (Based on data reported in *The Explosives Engineer*, vol. 1, 1976, p. 12.)

TABLE 3.11

x	1	2	3	4	5	6	7	8
$f(x)$	.02	.03	.05	.2	.4	.2	.07	?

(*a*) Find $f(8)$.
(*b*) Find the table for F.
(*c*) Use F to find the probability that a randomly selected bit can be used to drill between three and five holes inclusive.
(*d*) Find $P[X \leq 4]$ and $P[X < 4]$. Are these probabilities the same?
(*e*) Find $F(-3)$ and $F(10)$. *Hint:* Express these in terms of the probabilities that they represent and their values will become obvious.

9. Consider Example 1.2.1. Let X denote the number of computer systems operable at the time of the launch. Assume that the probability that each system is operable is .9.
   (*a*) Use the tree of Fig. 1.1 to find the density table.
   (*b*) There is a pattern to the probabilities in the density table. In particular

$$f(x) = k(x)(.9)^x(.1)^{3-x}$$

   where $k(x)$ gives the number of paths through the tree yielding a particular value for X. Use Exercise 17 of Chap. 1 to express $k(x)$ in terms of the number of computers available and the number operable.
   (*c*) Find the table for F.
   (*d*) Use F to find the probability that at least one system is operable at launch time.
   (*e*) Use F to find the probability that at most one system is operable at the time of the launch.

10. It is known that the probability of being able to log on to a computer from a remote terminal at any given time is .7. Let X denote the number of attempts that must be made to gain access to the computer.
   (*a*) Find the first four terms of the density table.
   (*b*) Find a closed form expression for $f(x)$.
   (*c*) Find $P[X = 6]$.
   (*d*) Find a closed form expression for $F(x)$.
   (*e*) Use F to find the probability that at most four attempts must be made to gain access to the computer.
   (*f*) Use F to find the probability that at least five attempts must be made to gain access to the computer.

In parts (*c*), (*d*), and (*e*) of each of the next two exercises, we point out the necessary and sufficient conditions for a function F to be a cumulative distribution function for a discrete random variable.

*11. Even though there is no closed-form expression for the cumulative distribution function of Exercise 7, we can rewrite it as follows:

$$F(x) = \begin{cases} 0 & x < 0 \\ .70 & 0 \le x < 1 \\ .90 & 1 \le x < 2 \\ .95 & 2 \le x < 3 \\ .98 & 3 \le x < 4 \\ .99 & 4 \le x < 5 \\ 1.00 & x \ge 5 \end{cases}$$

   (*a*) Draw the graph of this function. Recall from elementary calculus that graphs of this form are called step functions.
   (*b*) Is F a continuous function?
   (*c*) Is F a right continuous function?
   (*d*) What is $\lim_{x\to\infty} F(x)$? What is $\lim_{x\to-\infty} F(x)$?
   (*e*) Is F nondecreasing?

*12. (*a*) Express the cumulative distribution function F of Exercise 8 in the manner shown in Exercise 11 and draw the graph of F.
   (*b*) Is F a continuous function?
   (*c*) Is F right continuous?

(d) What is $\lim_{x \to \infty} F(x)$? What is $\lim_{x \to -\infty} F(x)$?
(e) Is F nondecreasing?

*13. State the conditions that are necessary and sufficient for a function F to be a cumulative distribution function for a discrete random variable.

Section 3.3

14. In an experiment to graft Florida sweet orange trees to the root of a sour orange variety, a series of five trials is conducted. Let X denote the number of grafts that fail. The density for X is given in Table 3.10.
(a) Find $E[X]$.
(b) Find μ_X.
(c) Find $E[X^2]$.
(d) Find Var X.
(e) Find σ_X^2.
(f) Find the standard deviation for X.
(g) What physical unit is associated with σ_X?

15. The density for X, the number of holes that can be drilled per bit while drilling into limestone is given in Table 3.11.
(a) Find $E[X]$ and $E[X^2]$.
(b) Find Var X and σ_X.
(c) What physical unit is associated with σ_X?

16. Use the density derived in Exercise 9, to find the expected value and variance for X, the number of computer systems operable at the time of the launch. Can you express $E[X]$ and Var X in terms of n, the number of systems available, and p, the probability that a given system will be operable?

*17. The probability p of being able to log on to a computer from a remote terminal at any given time is .7. Let X denote the number of attempts that must be made to gain access to the computer. Find $E[X]$. Can you express $E[X]$ in terms of p? *Hint:* The series $\sum_{x=1}^{\infty} x(.7)(.3)^{x-1} = E[X]$ is not geometric. To find $E[X]$, expand this series and the series $.3E[X]$. Subtract the two to form the series $.7E[X]$. Evaluate this *geometric* series and solve for $E[X]$.

*18. The probability that a cell will fuse in the presence of polyethylene glycol is $1/2$. Let Y denote the number of cells exposed to antigen-carrying lymphocytes to obtain the first fusion. Use the method of Exercise 17 to find $E[Y]$.

*19. Let X be a discrete random variable with density f. Let c be any real number. Show that
(a) $E[c] = c$. *Hint:* Remember that constants can be factored from summations and that $\sum_{\text{all } x} f(x) = 1$.
(b) $E[cX] = cE[X]$.

*20. Use the rules for expectation to verify that Var $c = 0$ and Var $cX = c^2$ Var X for any real number c. *Hint:* Var $c = E[c^2] - (E[c])^2$.

21. Let X and Y be independent random variables with $E[X] = 3$, $E[X^2] = 25$, $E[Y] = 10$ and $E[Y^2] = 164$.
(a) Find $E[3X + Y - 8]$.
(b) Find $E[2X - 3Y + 7]$.
(c) Find Var X.
(d) Find σ_X.
(e) Find Var Y.
(f) Find σ_Y.

(*g*) Find $\mathrm{Var}[3X + Y - 8]$.
(*h*) Find $\mathrm{Var}[2X - 3Y + 7]$.
(*i*) Find $E[(X - 3)/4]$ and $\mathrm{Var}[(X - 3)/4]$.
(*j*) Find $E[(Y - 10)/8]$ and $\mathrm{Var}[(Y - 10)/8]$.
(*k*) The results of parts (*i*) and (*j*) are not coincidental. Can you generalize and verify the conjecture suggested by these two exercises?

*22. Consider the function f defined by

$$f(x) = (1/2)2^{-|x|} \qquad x = \pm 1, \pm 2, \pm 3, \pm 4, \ldots$$

(*a*) Verify that this is the density for a discrete random variable X. *Hint:* Expand the series $\Sigma_{\text{all } x} f(x)$ for a few terms. A recognizable series will develop!
(*b*) Let $g(X) = (-1)^{|X|-1}[2^{|X|}/(2|X| - 1)]$. Show that $\Sigma_{\text{all } x} g(x)f(x) < \infty$. *Hint:* Expand the series for a few terms. You will obtain an alternating series that can be shown to converge.
(*c*) Show that $\Sigma_{\text{all } x}|g(x)|f(x)$ does not converge. This will show that $E[g(X)]$ does not exist. *Hint:* Expand the series for a few terms. You will obtain a series that is term by term larger than the diverging harmonic type series $1/3\Sigma_{x=1}^{\infty} 1/x$.

*23. (*An application to sort algorithms.*) In studying various sort algorithms in computer science, it is of interest to compare their efficiency by estimating the average number of interchanges needed to sort random arrays of various sizes. It is also of interest to compare these estimated averages to the "ideal" average where by "ideal" we mean the expected minimum number of interchanges needed to sort the array. In this exercise you will derive this ideal average. "A Note on the Minimum Number of Interchanges Needed to Sort a Random Array," T. McMillan, I. Liss, and J. Milton, Radford University, Radford, Virginia.
(*a*) Consider a random array of length n. When the positions of exactly two elements of the array are exchanged, we say that an "interchange" has taken place. Let X_n denote the minimum number of interchanges necessary to sort an array of size n. Note that

$$X_n = X_{n-1} + I$$

where $I = 0$ if the last element of the array is in the correct position and $I = 1$ otherwise. Argue that $P[I = 0] = 1/n$ and $P[I = 1] = 1 - (1/n)$.
(*b*) Show that

$$E[I] = 1 - \frac{1}{n}$$

(*c*) Argue that

$$\begin{aligned}
E[X_n] &= E[X_{n-1}] + 1 - \frac{1}{n} \\
E[X_{n-1}] &= E[X_{n-2}] + 1 - \frac{1}{n-1} \\
E[X_{n-2}] &= E[X_{n-3}] + 1 - \frac{1}{n-2} \\
&\vdots \\
E[X_3] &= E[X_2] + 1 - \frac{1}{3} \\
E[X_2] &= E[X_1] + 1 - \frac{1}{2} \\
E[X_1] &= 0
\end{aligned}$$

(*d*) Use a recursive argument to show that

$$E[X_n] = (n-1) - \sum_{i=2}^{n} \frac{1}{i}$$

(*e*) Illustrate the expression given in part (*d*) by finding $E[X_5]$.

(*f*) Elementary calculus can be used to approximate $E[X_n]$ by noting that

$$\sum_{i=2}^{n} \frac{1}{i} \doteq \int_{1.5}^{n+.5} \frac{1}{t}\,dt$$

Use this idea to approximate $E[X_5]$ and compare the result to the exact solution found in part (*e*).

(*g*) A random digit generator is used to generate sets of 100 different three-digit numbers lying between 0 and 1. What is the ideal average number of interchanges needed to sort such an array?

Section 3.4

24. The probability that a wildcat well will be productive is 1/13. Assume that a group is drilling wells in various parts of the country so that the status of one well has no bearing on that of any other. Let X denote the number of wells drilled to obtain the first strike.

(*a*) Verify that X is geometric and identify the value of the parameter p.
(*b*) What is the exact expression for the density for X?
(*c*) What is the exact expression for the moment generating function for X?
(*d*) What are the numerical values of $E[X]$, $E[X^2]$, σ^2, and σ?

25. The zinc-phosphate coating on the threads of steel tubes used in oil and gas wells is critical to their performance. To monitor the coating process, an uncoated metal sample with known outside area is weighed and treated along with the lot of tubing. This sample is then stripped and reweighed. From this it is possible to determine whether or not the proper amount of coating was applied to the tubing. Assume that the probability that a given lot is unacceptable is .05. Let X denote the number of runs conducted to produce an unacceptable lot. Assume that the runs are independent in the sense that the outcome of one run has no effect on that of any other. (Based on a report in *American Machinist*, November 1982, p. 81.)

(*a*) Verify that X is geometric. What is "success" in this experiment? What is the numerical value of p?
(*b*) What is the exact expression for the density for X?
(*c*) What is the exact expression for the moment generating function for X?
(*d*) What are the numerical values of $E[X]$, $E[X^2]$, σ^2, and σ?

26. A system used to read electric meters automatically requires the use of a 128-bit computer message. Occasionally random interference causes a digit reversal resulting in a transmission error. Assume that the probability of a digit reversal for each bit is 1/1000. Let X denote the number of transmission errors per 128-bit message sent. Is X geometric? If not, what geometric property fails?

27. Verify that the random variable X of Exercise 17 is geometric. Use Theorem 3.4.3 to find $E[X]$ and compare your answer to that obtained in Exercise 17.

28. Verify that the random variable Y of Exercise 18 is geometric. Use Theorem 3.4.3 to find $E[Y]$ and compare your answer to that obtained in Exercise 18.

29. Consider the random variable X whose density is given by

$$f(x) = \frac{(x-3)^2}{5} \qquad x = 3, 4, 5$$

(*a*) Verify that this function is a density for a discrete random variable.
(*b*) Find $E[X]$ directly. That is, evaluate $\Sigma_{\text{all } x} x f(x)$.
(*c*) Find the moment generating function for X.
(*d*) Use the moment generating function to find $E[X]$ thus verifying your answer to part (*b*) of this exercise.
(*e*) Find $E[X^2]$ directly. That is, evaluate $\Sigma_{\text{all } x} x^2 f(x)$.
(*f*) Use the moment generating function to find $E[X^2]$ thus verifying your answer to part (*e*) of this exercise.
(*g*) Find σ^2 and σ.

30. A discrete random variable has moment generating function

$$m_X(t) = e^{2(e^t - 1)}$$

(*a*) Find $E[X]$.
(*b*) Find $E[X^2]$.
(*c*) Find σ^2 and σ.

*31. Let X have a geometric distribution with parameter p.
(*a*) Show that the probability that X is odd is $p/(1-q^2)$ where $q = 1 - p$. *Hint:* If x is odd, then x can be expressed in the form $x = 2m - 1$ for $m = 1, 2, 3, \ldots$.
(*b*) Show that the probability that X is odd is never $1/2$ regardless of the value chosen for p.

32. (*Discrete uniform distribution.*) A discrete random variable is said to be *uniformly* distributed if it assumes a finite number of values with each value occurring with the same probability. If we consider the generation of a single random digit then Y, the number generated, is uniformly distributed with each possible digit occurring with probability $1/10$. In general, the density for a uniformly distributed random variable is given by

$$f(x) = 1/n \qquad n \text{ a positive integer}$$
$$x = x_1, x_2, x_3, \ldots, x_n$$

(*a*) Find the moment generating function for a discrete uniform random variable.
(*b*) Use the moment generating function to find $E[X]$, $E[X^2]$, and σ^2.
(*c*) Find the mean and variance for the random variable Y, the number obtained when a random digit generator is activated once. *Hint:* The sum of the first n positive integers is $n(n+1)/2$; the sum of the squares of the first n positive integers is $n(n+1)(2n+1)/6$.

33. Let the density for X be given by

$$f(x) = ce^{-x} \qquad x = 1, 2, 3, \ldots$$

(a) Find the value of c that makes this a density.
(b) Find the moment generating function for X.
(c) Use $m_X(t)$ to find $E[X]$.

Section 3.5

34. Let X be binomial with parameters $n = 15$ and $p = .2$.
(a) Find the expression for the density for X.
(b) Find the expression for the moment generating function for X.
(c) Find $E[X]$ and Var X.
(d) Find $E[X]$, $E[X^2]$, and Var X using the moment generating function thus verifying your answer to part (c) of this exercise.
(e) Find $P[X \leq 1]$ by evaluating the density directly. Compare your answer to that given in Table I of App. A.
(f) Use Table I of App. A to find each of these probabilities.

$$P[X \leq 5] \qquad P[X \geq 3]$$

$$P[X < 5] \qquad F(9)$$

$$P[2 \leq X \leq 7] \qquad F(20)$$

$$P[2 \leq X < 7] \qquad P[X = 10]$$

35. Albino rats used to study the hormonal regulation of a metabolic pathway are injected with a drug that inhibits body synthesis of protein. The probability that a rat will die from the drug before the experiment is over is .2. If 10 animals are treated with the drug, how many are expected to die before the experiment ends? What is the probability that at least eight will survive? Would you be surprised if at least five died during the course of the experiment? Explain, based on the probability of this occurring.

36. Consider Example 1.2.1. The random variable X is the number of computer systems operable at the time of a space launch. The systems are assumed to operate independently. Each is operable with probability .9.
(a) Argue that X is binomial and find its density. Compare your answer to that obtained in Exercise 9(b).
(b) Find $E[X]$ and Var X.

37. In humans, geneticists have identified two sex chromosomes, R and Y. Every individual has an R chromosome, and the presence of a Y chromosome distinguishes the individual as male. Thus the two sexes are characterized as RR (female) and RY (male). Color blindness is caused by a recessive allele on the R chromosome, which we denote by r. The Y chromosome has no bearing on color blindness. Thus relative to color blindness, there are three genotypes for females and two for males:

Female	**Male**
RR (normal)	RY (normal)
Rr (carrier)	rY (color-blind)
rr (color-blind)	

A child inherits one sex chromosome randomly from each parent.

(*a*) A carrier of color blindness parents a child with a normal male. Construct a tree to represent the possible genotypes for the child. Use the tree to find the probability that a given child will be a color-blind male.

(*b*) If the couple has five children, what is the expected number of color-blind males? What is the probability that three or more will be color-blind males?

38. In scanning electron microscopy photography, a specimen is placed in a vacuum chamber and scanned by an electron beam. Secondary electrons emitted from the specimen are collected by a detector and an image is displayed on a cathode ray tube. This image is photographed. In the past a 4×5 in camera has been used. It is thought that a 35-mm camera can obtain the same clarity. This type of camera is faster and more economical than the 4×5-in variety. (Based on a report entitled "Adaptation of a Thirty-five Millimeter Photographic System for a Scanning Electron Microscope," E. A. Lawton, *Biological Photography*, volume 50, no. 3, July, 1982, p. 65.)

(*a*) Photographs of 15 specimens are made using each camera system. These unmarked photographs are judged for clarity by an impartial judge. The judge is asked to select the better of the two photographs from each pair. Let X denote the number selected taken by a 35-mm camera. If there is really no difference in clarity and the judge is randomly selecting photographs, what is the expected value of X?

(*b*) Would you be surprised if the judge selected 12 or more photographs taken by the 35-mm camera? Explain, based on the probability involved.

(*c*) If $X \geq 12$, do you think that there is reason to suspect that the judge is not selecting the photographs at random?

39. It has been found that 80% of all printers used on home computers operate correctly at the time of installation. The rest require some adjustment. A particular dealer sells 10 units during a given month.

(*a*) Find the probability that at least nine of the printers operate correctly upon installation.

(*b*) Consider five months in which 10 units are sold per month. What is the probability that at least nine units operate correctly in each of the 5 months?

40. It is possible for a computer to pick up an erroneous signal that does not show up as an error on the screen. The error is called a silent error. A particular terminal is defective and when using the system word processor it introduces a silent paging error with probability .1. The word processor is used 20 times during a given week.

(*a*) Find the probability that no silent paging errors occur.

(*b*) Find the probability that at least one such error occurs.

(*c*) Would it be unusual for more than four such errors to occur? Explain, based on the probability involved.

41. (*a*) Find the moment generating function for a binomial random variable with parameters n and p. *Hint:* Let

$$\binom{n}{x} e^{tx} p^x (1-p)^{n-x} = \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

and apply the binomial theorem.

(*b*) Use $m_X(t)$ to show that $E[X] = np$.

(*c*) Use $m_X(t)$ to show that $E[X^2] = n^2p^2 - np^2 + np$.

(*d*) Show that $\text{Var}\, X = npq$ where $q = 1 - p$.

⋆**42.** Find the mean value for a binomial random variable with parameters n and p from the definition. That is, evaluate

$$\sum_{x=0}^{n} x\binom{n}{x} p^x (1-p)^{n-x}$$

Hint:

$$\sum_{x=0}^{n} x\binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x\binom{n}{x} p^x (1-p)^{n-x}$$

Now let $z = x - 1$ and evaluate

$$\sum_{z=0}^{n-1} (z+1)\binom{n}{z+1} p^{z+1} (1-p)^{n-(z+1)}$$

43. (*Point binomial or Bernoulli distribution.*) Assume that an experiment is conducted and that the outcome is considered to be either a success or a failure. Let p denote the probability of success. Define X to be 1 if the experiment is a success and 0 if it is a failure. X is said to have a *point binomial* or a *Bernoulli* distribution with parameter p.

(*a*) Argue that X is a binomial random variable with $n = 1$.

(*b*) Find the density for X.

(*c*) Find the moment generating function for X.

(*d*) Find the mean and variance for X.

(*e*) In DNA replication errors can occur that are chemically induced. Some of these errors are "silent" in that they do not lead to an observable mutation. Growing bacteria are exposed to a chemical that has probability .14 of inducing an observable error. Let X be 1 if an observable mutation results and let X be 0 otherwise. Find $E[X]$.

44. A binomial random variable has mean 5 and variance 4. Find the values of n and p that characterize the distribution of this random variable.

Section 3.6

***45.** In this exercise you will show that the density for the negative binomial distribution sums to 1.

(*a*) Show that

$$\sum_{x=r}^{\infty} \binom{x-1}{r-1} q^{x-r} p^r = \sum_{z=0}^{\infty} \binom{z+r-1}{r-1} q^z p^r$$

where $q = 1 - p$. *Hint:* Let $z = x - r$.

(*b*) Show that the Taylor series expansion for $h(q) = 1/(1-q)^r$ about 0 is given by

$$1 + rq + \frac{r(r+1)q^2}{2} + \frac{r(r+1)(r+2)q^3}{3!} + \cdots = 1/(1-q)^r$$

(*c*) Show that

$$\sum_{z=0}^{\infty} \binom{z+r-1}{r-1} q^z = 1/(1-q)^r$$

(*d*) Show that

$$\sum_{z=0}^{\infty} \binom{z+r-1}{r-1} q^z p^r = 1$$

Hint: Factor the term p^r from the expression and use part (c) with $q = 1 - p$.

*46. In this exercise you will derive the moment generating function for the negative binomial distribution with parameters r and p.

(*a*) Show that

$$m_X(t) = \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} q^{x-r} p^r = (pe^t)^r \sum_{z=0}^{\infty} \binom{z+r-1}{r-1} (qe^t)^z$$

Hint: Let $z = x - r$.

(*b*) Use the idea given in Exercise 45(c) to show that

$$\sum_{z=0}^{\infty} \binom{z+r+1}{r-1} (qe^t)^z = 1/(1-qe^t)^r$$

(*c*) Show that

$$m_X(t) = \frac{(pe^t)^r}{(1-qe^t)^r}$$

*47. Use the moment generating function to show that the mean of a negative binomial distribution with parameters r and p is r/p.

*48. Use the moment generating function to show that $E[X^2] = (r^2 + rq)/p^2$ and that $\text{Var } X = rq/p^2$ for the negative binomial distribution with parameters r and p.

*49. Show that the geometric distribution is a special case of the negative binomial distribution with $r = 1$. Find the mean and variance of a geometric random variable with parameter p using Exercises 47 and 48. Compare your answer with the results of Theorem 3.4.3.

*50. A vaccine for desensitizing patients to bee strings is to be packed with three vials in each box. Each vial is checked for strength before packing. The probability that a vial meets specifications is .9. Let X denote the number of vials that must be checked to fill a box. Find the density for X and its mean and variance. Would you be surprised if seven or more vials have to be tested to find three that meet specifications? Explain, based on the probability of this occurrence.

*51. Some characteristics in animals are said to be sex-influenced. For example, the production of horns in sheep is governed by a pair of alleles, H and h. The allele H for the production of horns is dominant in males but recessive in females. The allele h for hornlessness is dominant in females and recessive in males. Thus, given a heterozygous male (Hh) and a heterozygous female (Hh), the male will have horns but the female will be hornless. Assume that two such animals mate and the offspring is just as likely to be male as female. The lamb inherits one gene for horns randomly from each parent. Use a tree diagram to show that the probability that a lamb will be a hornless female is 3/8. Find the average number of lambs born to obtain the second hornless female. Would you be surprised if at most five lambs were born to obtain the second hornless female? Explain.

Section 3.7

52. Suppose that X is hypergeometric with $N = 20$, $r = 17$, and $n = 5$. What are the possible values for X? What is $E[X]$ and Var X?

53. Suppose that X is hypergeometric with $N = 20$, $r = 3$, and $n = 5$. What are the possible values for X? What is $E[X]$ and Var X?

54. Suppose that X is hypergeometric with $N = 20$, $r = 10$, and $n = 5$. What are the possible values for X? What is $E[X]$ and Var X?

55. Twenty microprocessor chips are in stock. Three have etching errors that cannot be detected by the naked eye. Five chips are selected and installed in field equipment.
(*a*) Find the density for X, the number of chips selected that have etching errors.
(*b*) Find $E[X]$ and Var X.
(*c*) Find the probability that no chips with etching errors will be selected.
(*d*) Find the probability that at least one chip with an etching error will be chosen.

56. Production line workers assemble 15 automobiles per hour. During a given hour, four are produced with improperly fitted doors. Three automobiles are selected at random and inspected. Let X denote the number inspected that have improperly fitted doors.
(*a*) Find the density for X.
(*b*) Find $E[X]$ and Var X.
(*c*) Find the probability that at most one will be found with improperly fitted doors.

57. A distributor of computer software wants to obtain some customer feedback concerning its newest package. Three thousand customers have purchased the package. Assume that 600 of these customers are dissatisfied with the product. Twenty customers are randomly sampled and questioned about the package. Let X denote the number of dissatisfied customers sampled.
(*a*) Find the density for X.
(*b*) Find $E[X]$ and Var X.
(*c*) Set up the calculations needed to find $P[X \leq 3]$.
(*d*) Use the binomial tables to approximate $P[X \leq 3]$.

58. A random telephone poll is conducted to ascertain public opinion concerning the construction of a nuclear power plant in a particular community. Assume that there are 150,000 numbers listed for private individuals and that 90,000 of these would elicit a negative response if contacted. Let X denote the number of negative responses obtained in 15 calls.
(*a*) Find the density for X.
(*b*) Find $E[X]$ and Var X.
(*c*) Set up the calculations needed to find $P[X \geq 6]$.
(*d*) Use the binomial tables to approximate $P[X \geq 6]$.

Section 3.8

59. Let X be a Poisson random variable with parameter $k = 10$.
(*a*) Find $E[X]$.
(*b*) Find Var X.
(*c*) Find σ_X.
(*d*) Find the expression for the density for X.
(*e*) Find $P[X \leq 4]$.
(*f*) Find $P[X < 4]$.

(*g*) Find $P[X = 4]$.
(*h*) Find $P[X \geq 4]$.
(*i*) Find $P[4 \leq X \leq 9]$.

60. A particular nuclear plant releases a detectable amount of radioactive gases twice a month on the average. Find the probability that there will be at most four such emissions during a month. What is the expected number of emissions during a three-month period? If, in fact, 12 or more emissions are detected during a three-month period, do you think that there is a reason to suspect the reported average figure of twice a month? Explain, on the basis of the probability involved.

61. Geophysicists determine the age of a zircon by counting the number of uranium fission tracks on a polished surface. A particular zircon is of such an age that the average number of tracks per square centimeter is five. What is the probability that a 2-cm^2 sample of this zircon will reveal at most three tracks thus leading to an underestimation of the age of the material?

62. California is hit by approximately 500 earthquakes that are large enough to be felt every year. However, those of destructive magnitude occur on the average once every year. Find the probability that California will experience at least one earthquake of this magnitude during a six-month period. Would it be unusual to have three or more earthquakes of destructive magnitude in a six-month period? Explain, based on the probability of this occurring. (Based on data presented in *Earthquake Country* by Robert Iacopi.)

63. Load-bearing structures in underground mines are often required to carry additional loads while mining operations are in progress. As the structures adjust to this new weight small-scale displacements take place that result in the release of seismic and acoustic energy called "rock noise." This energy can be detected using special geophysical equipment. Assume that in a particular mine the average number of rock noises recorded during normal activity is three per hour. Would you consider it unusual if more than 10 were detected in a two-hour period? Explain, based on the probability involved. (Based on "A multichannel rock noise monitoring system," T. Gowd and M. S. Rao, *Journal of Mines, Metals and Fuels*, September, 1981, pp. 288–290.)

64. A burr is a thin ridge or rough area that occurs when shaping a metal part. These must be removed by hand or by means of some newer method such as water jets, thermal energy, or electrochemical processing before the part can be used. Assume that a part used in automatic transmissions typically averages two burrs each. What is the probability that the total number of burrs found on seven randomly selected parts will be at most four? (Based on "Advances in Deburring" by B. Hignett, *Production Engineering*, December, 1982, pp. 44–47.)

65. Cast iron is an alloy composed primarily of iron together with smaller amounts of other elements including carbon, silicon, sulfur, and phosphorus. The carbon occurs as graphite, which is soft, or iron carbide, which is very hard and brittle. The type of cast iron produced is determined by the amount and distribution of carbon in the iron. Five types of cast iron are identifiable. These are gray, compacted graphite, ductile, malleable, and white. In malleable cast iron the carbon is present as discrete graphite particles. Assume that in a particular casting these particles average 20 per square inch. Would it be unusual to see a $\frac{1}{4}$-in^2 area of this casting with fewer than two graphite particles? Explain, based on the probability involved. (Based on "Space Age Metal: Cast Iron," J. Lalich, *Mines Magazine*, February, 1982, pp. 2–6.)

66. A Poisson random variable is such that it assumes the values 0 and 1 with equal probability. Find the value of the Poisson parameter k for this variable.

67. Prove Theorem 3.8.1. *Hint:* Note that

$$m_X(t) = E[e^{tX}] = \sum_{x=0}^{\infty} \frac{e^{tx}e^{-k}k^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-k}(ke^t)^x}{x!}$$

and use the Maclaurin series.

68. Let X be binomial with $n = 20$ and $p = .05$. Find $P[X = 0]$ using the binomial density and compare your answer to that obtained using the Poisson approximation to this probability. Do you think that the error in the approximation is large?

69. In *Escherichia coli*, a bacterium often found in the human digestive tract, 1 cell in every 10^9 will mutate from streptomycin sensitivity to streptomycin resistance. This mutation can cause the individual involved to become resistant to the antibiotic streptomycin. In observing 2 billion (2×10^9) such cells, what is the probability that none will mutate? What is the probability that at least one will mutate?

70. The spontaneous flipping of a bit stored in a computer memory is called a "soft fail." Soft fails are rare, averaging only one per million hours per chip. However, the probability of a soft fail is increased when the chip is exposed to α particles (helium nuclei) which occur naturally in the environment. Assume that the probability of a soft fail under these conditions is $1/1000$. If a chip containing 6000 bits is exposed to α particles, what is the probability that there will be at least one soft fail? Would you be surprised if there were more than five soft fails? Explain, based on the probability of this occurring. (*McGraw-Hill Yearbook of Science and Technology*, 1981, p. 142.)

Section 3.9

71. Use Table II of App. A to simulate the arrival and departure of planes to the airport described in Example 3.8.2 for 10 more one-minute periods. Based on these data, approximate the average number of planes on the ground at a given time by finding the arithmetic average of the values of Z simulated in the experiment.

72. Consider the random variable X, the number of runs conducted to produce an unacceptable lot when coating steel tubes (see Exercise 25.) X is geometric with $p = .05$. Divide the 100 possible two-digit numbers into two categories with numbers 00–04 denoting the production of an unacceptable lot and the remaining numbers denoting the production of an acceptable lot. Simulate the experiment of producing lots until an unacceptable one is obtained 10 times. Record the value obtained for X in each simulation. Based on these data, approximate the average value of X. Does your approximate value lie close to the theoretical mean value of 20? If not, run the simulation 10 more times. Is the arithmetic average of your observed values for X closer to 20 this time?

REVIEW EXERCISES

73. A large microprocessor chip contains multiple copies of circuits. If a circuit fails, the chip knows it and knows how to select the proper logic to repair itself. The average number of defects per chip is 300. What is the probability that 10 or fewer defects will be found in a randomly selected region that comprises 5% of the total surface area? What is the probability that more than 10 defects are found? ("Self-Repairing Chips," *Datamation*, May, 1983, p. 68.)

74. When a program is submitted to the computer in a time-sharing system, it is processed on a space-available basis. Past experience shows that a program submitted to one such system is accepted for processing within one minute with probability .25. Assume that during the course of a day five programs are submitted with enough time between submissions to ensure independence. Let X denote the number of programs accepted for processing within one minute.
   (*a*) Find $E[X]$ and Var X.
   (*b*) Find the probability that none of these programs will be accepted for processing within one minute.
   (*c*) Five programs are submitted on each of two consecutive days. What is the probability that no programs will be accepted for processing within one minute during this two-day period?

75. A new type of brake lining is being studied. It is thought that the lining will last for at least 70,000 miles on 90% of the cars in which it is used. Laboratory trials are conducted to simulate the driving experience of 100 cars in which this lining is used. Let X denote the number of cars whose brakes must be relined before the 70,000 mile mark.
   (*a*) What is the distribution of X? What is $E[X]$?
   (*b*) What distribution can be used to approximate probabilities for X?
   (*c*) Suppose that we agree that the 90% figure is too high if 17 or more of the 100 cars require a relinement prior to the 70,000-mile mark. What is the probability that we will come to this conclusion by chance even though the 90% figure is correct?

76. A bank of guns fires on a target one after the other. Each has probability $1/4$ of hitting the target on a given shot. Find the probability that the second hit comes before the seventh gun fires.

77. In a video game, the player attempts to capture a treasure lying behind one of five doors. The location of the treasure varies randomly in such a way that at any given time it is just as likely to be behind one door as any other. When the player knocks on a given door, the treasure is his if it lies behind that door. Otherwise, he must return to his original starting point and approach the doors through a dangerous maze again. Once the treasure is captured the game ends. Let X denote the number of trials needed to capture the treasure. Find the average number of trials needed to capture the treasure. Find $P[X \le 3]$. Find $P[X > 3]$.

78. An automobile repair shop has 10 rebuilt transmissions in stock. Three are not in correct working order and have an internal defect that will cause trouble within the first 1000 miles of operation. Four of these transmissions are randomly selected and installed in customers' cars. Find the probability that no defective transmissions are installed. Find the probability that exactly one defective transmission is installed.

79. A computer terminal can pick up an erroneous signal from the keyboard that does not show up on the screen. This creates a silent error that is difficult to detect. Assume that, for a particular keyboard, the probability that this will occur per entry is $1/1000$. In 12,000 entries find the probability that no silent errors occur. Find the probability of at least one silent error.

80. It is thought that one of every ten cars on the road have speedometers that are miscalibrated to the extent that they read at least five miles per hour low. During the course of a day, 15 drivers are stopped and charged with exceeding the speed limit by at least five miles per hour. Would you be surprised to find that at least five of the cars involved have miscalibrated speedometers? Explain, based on the probability of observing a result this unusual by chance.

81. Let

$$f(x) = \frac{x^2}{14} \qquad x = 1, 2, 3$$

(*a*) Show that f is the density for a discrete random variable.
(*b*) Find $E[X]$ and $E[X^2]$ from the definition of these terms.
(*c*) Find $m_X(t)$.
(*d*) Use $m_X(t)$ to verify your answers to part (*b*).
(*e*) Find Var X and σ.

CHAPTER 4

CONTINUOUS DISTRIBUTIONS

In Chap. 3 we learned to distinguish a discrete random variable from one that is not discrete. In this chapter we consider a large class of nondiscrete random variables. In particular, we consider random variables that are called "continuous." We first study the general properties of variables of the continuous type and then present some important families of continuous random variables.

4.1 CONTINUOUS DENSITIES

In Chap. 3 we considered the random variable T, the time of the peak demand for electricity at a particular power plant. We agreed that this random variable is not discrete since, "a priori"—before the fact, we cannot limit the set of possible values for T to some finite or countably infinite collection of times. Time is measured continuously and T can conceivably assume any value in the time interval $[0, 24)$ where 0 denotes 12 midnight one day and 24 denotes 12 midnight the next day. Furthermore, if we ask *before* the day begins, what is the probability that the peak demand will occur exactly 12.013 278 650 931 271? The answer is 0. It is virtually impossible for the peak load to occur at this split second in time, not the slightest bit earlier or later. These two properties, possible values occurring as intervals and the a priori probability of assuming any specific value being 0, are the characteristics that identify a random variable as being continuous. This leads us to our next definition.

Definition 4.1.1 (Continuous random variable). A random variable is continuous if it can assume any value in some interval or intervals of real numbers and the probability that it assumes any specific value is 0.

Note that the statement that the probability that a continuous random variable assumes any specific value is 0 is essential to the definition. Discrete variables have no such restriction. For this reason, we calculate probabilities in the continuous case differently than we do in the discrete case. In the discrete case, we defined a function f, called the density, which enabled us to compute probabilities associated with the random variable X. This function is given by

$$f(x) = P[X = x] \qquad x \text{ real}$$

This definition cannot be used in the continuous case because $P[X = x]$ is always 0. However, we do need a function that will enable us to compute probabilities associated with a continuous random variable. Such a function is also called a density.

Definition 4.1.2 (Continuous density). Let X be a continuous random variable. A function f such that

1. $f(x) \geq 0 \qquad$ for x real
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $P[a \leq X \leq b] = \int_a^b f(x)\,dx \qquad$ for a and b real

is called a density for X.

Although this definition may look arbitrary at first glance, it is not. Note that, as in the discrete case, f is defined over the entire real line and is nonnegative. Recall from elementary calculus that integration is the natural extension of summation in the sense that the integral is the limit of a sequence of Riemann sums. In the discrete case, we require that $\sum_{\text{all } x} f(x) = 1$. The natural extension of this requirement to the continuous case is property 2, namely, that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

In the discrete case, we find the probability that X assumes a value in some set A by summing $f(x)$ over all values of x in A. That is,

$$P[X \in A] = \sum_{x \in A} f(x).$$

In the continuous case, we shall be interested in finding the probability that X assumes values in some interval $[a, b]$. Replacing A by $[a, b]$ and substituting integration for summation in the previous expression suggests property 3 of Definition 4.1.2. That is,

$$P[a \leq X \leq b] = \int_a^b f(x)\,dx$$

It is evident that the term "density" in the continuous case is just an extension of the ideas presented in the discrete case with summation being replaced by integration. This is an important notion as it will allow us to define the concept of expected value in the continuous case quite naturally.

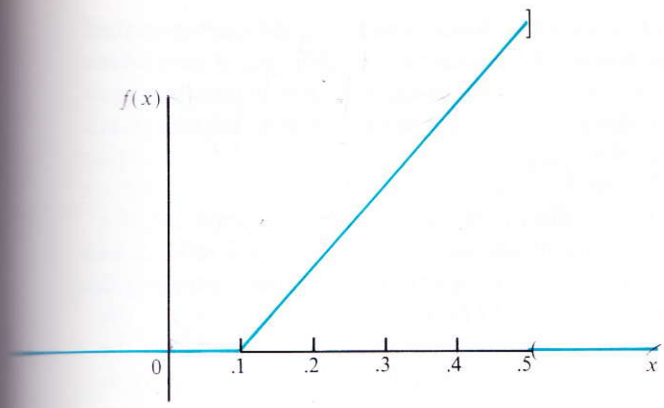


FIGURE 4.1
Graph of
$$f(x) = \begin{cases} 12.5x - 1.25 & .1 \le x \le .5 \\ 0 & \text{elsewhere} \end{cases}$$

Example 4.1.1. The lead concentration in gasoline currently ranges from .1 to .5 grams per liter. What is the probability that the lead concentration in a randomly selected liter of gasoline will lie between .2 and .3 grams inclusive? To answer this question, we need a density, f, for the random variable X, the number of grams of lead per liter of gasoline. Consider the function

$$f(x) = \begin{cases} 12.5x - 1.25 & .1 \le x \le .5 \\ 0 & \text{elsewhere} \end{cases}$$

The graph of f is shown in Fig. 4.1. The function is nonnegative. Furthermore,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_{.1}^{.5} (12.5x - 1.25)\,dx \\
&= \frac{12.5x^2}{2} - 1.25x \Big|_{.1}^{.5} \\
&= \left[\frac{12.5(.5)^2}{2} - 1.25(.5)\right] - \left[\frac{12.5(.1)^2}{2} - 1.25(.1)\right] \\
&= .9375 - (-.0625) = 1
\end{aligned}$$

Thus f satisfies properties 1 and 2 of Definition 4.1.2. Property 3 allows us to use f to find the desired probability. In particular

$$\begin{aligned}
P[.2 \le X \le .3] &= \int_{.2}^{.3} f(x)\,dx \\
&= \int_{.2}^{.3} (12.5x - 1.25)\,dx \\
&= \frac{12.5x^2}{2} - 1.25x \Big|_{.2}^{.3} \\
&= \left[\frac{12.5(.3)^2}{2} - 1.25(.3)\right] - \left[\frac{12.5(.2)^2}{2} - 1.25(.2)\right] \\
&= .1875
\end{aligned}$$

(Based on data reported in *Petroleum Review*, August, 1982, p. 45.)

There are several important points to be made concerning the density in the continuous case. First, we shall follow the convention of defining f only over intervals for which $f(x)$ may be nonzero. For values of x not explicitly mentioned, $f(x)$ is assumed to be 0. In Example 4.1.1, we could have written f as

$$f(x) = 12.5x - 1.25 \qquad .1 \le x \le .5$$

with the understanding that $f(x) = 0$ elsewhere. Second, since the integral of a nonnegative function can be thought of as an area, properties 2 and 3 of Definition 4.1.2 can be expressed in terms of areas. In particular, property 2 requires that *the total area under the graph of f be 1*. Property 3 implies that the probability that the variable assumes a value between two points a and b is the *area under the graph of f between $x = a$ and $x = b$*. These ideas as they apply to Example 4.1.1 are demonstrated in Figs. 4.2(a) and (b), respectively. Third, since $P[X = a] = P[X = b] = 0$ in the continuous case

$$P[a \le X \le b] = P[a \le X < b] = P[a < X \le b] = P[a < X < b].$$

In Example 4.1.1, the probability that the lead concentration in a liter of gasoline lies between .2 and .3 grams inclusive, $P[.2 \le X \le .3]$, is the same as $P[.2 < X < .3]$, the probability that it lies strictly between .2 and .3 grams. See Fig. 4.2(c). Fourth, properties 1 and 2 of Definition 4.1.2 are necessary and sufficient conditions for a function to be a density for a continuous random variable X.

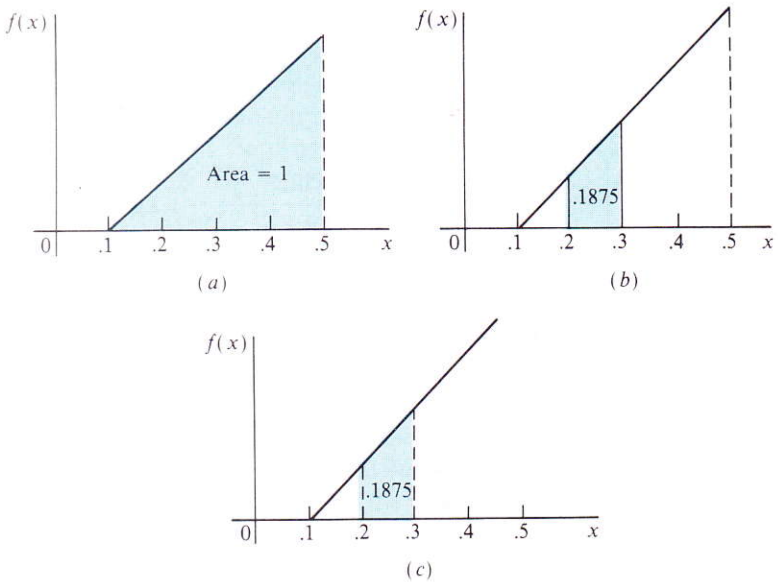


FIGURE 4.2
(a) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ implies that the total area under the graph of f is 1. (b) $P[.2 \le X \le .3] = \int_{.2}^{.3}(12.5x - 1.25)\,dx = .1875$ implies that the area under the graph of f between $x = .2$ and $x = .3$ is .1875. (c) $P[.2 < X < .3] = P[.2 \le X \le .3] = .1875$.

However, the density chosen for X can't be just any function satisfying these conditions. It should be a function that assigns reasonable probabilities to events via property 3 of Definition 4.1.2. Whether or not the function f given in Example 4.1.1 satisfies this criteria is debatable. It was chosen for illustrative purposes only. Finding an appropriate density is not always easy. Some methods for helping in the selection of a density are discussed in Chap. 6.

The idea of a cumulative distribution function in the continuous case is useful. It is defined exactly as in the discrete case although it is found using integration rather than summation.

Definition 4.1.3 (Cumulative distribution—continuous). Let X be continuous with density f. The cumulative distribution function for X, denoted by F, is defined by

$$F(x) = P[X \leq x] \qquad x \text{ real}$$

To find $F(x)$ for a specific real number x, we integrate the density over all real numbers that are less than or equal to x. Computationally

$$P[X \leq x] = F(x) = \int_{-\infty}^{x} f(t)\, dt \qquad x \text{ real}$$

Graphically, this probability corresponds to the area under the graph of the density to the left of and including the point x. Exercise 14 points out the mathematical properties of F.

Example 4.1.2. The density for the random variable X, the lead content in a liter of gasoline is

$$f(x) = 12.5x - 1.25 \qquad .1 \leq x \leq .5$$

The cumulative distribution function for X is

$$P[X \leq x] = F(x) = \int_{-\infty}^{x} f(t)\, dt$$

For $x < .1$, this integral has value 0 since for these values of x, $f(t)$ is itself 0. For $.1 \leq x \leq .5$

$$\begin{aligned} F(x) = \int_{-\infty}^{x} f(t)\, dt &= \int_{.1}^{x} (12.5t - 1.25)\, dt \\ &= \frac{12.5t^2}{2} - 1.25t \Big|_{.1}^{x} \\ &= 6.25x^2 - 1.25x + .0625 \end{aligned}$$

For $x > .5$, the integral has value 1 since, for these values of x, we have integrated the density over its entire set of possible values. Summarizing, F is given by

$$F(x) = \begin{cases} 0 & x < .1 \\ 6.25x^2 - 1.25x + .0625 & .1 \leq x \leq .5 \\ 1 & x > .5 \end{cases}$$

What is the probability that the lead concentration in a randomly selected liter of

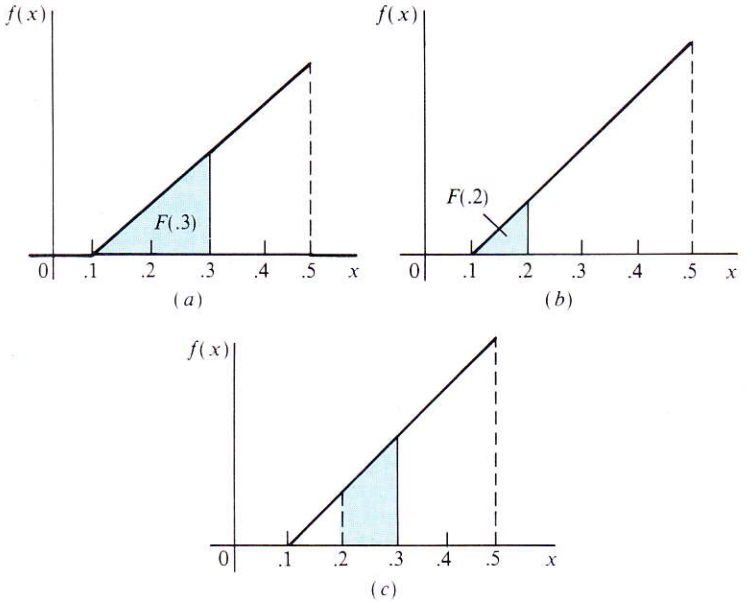


FIGURE 4.3
(a) $F(.3) = P[X \le .3]$. (b) $F(.2) = P[X \le .2]$. (c) $F(.3) - F(.2) = P[.2 \le X \le .3]$.

gasoline will lie between .2 and .3 grams per liter? To answer this question, we rewrite it in terms of the cumulative distribution

$$\begin{aligned} P[.2 \le X \le .3] &= P[X \le .3] - P[X < .2] \\ &= P[X \le .3] - P[X \le .2] \qquad (X \text{ is continuous}) \\ &= F(.3) - F(.2) \end{aligned}$$

By substitution

$$F(.3) = 6.25(.3)^2 - 1.25(.3) + .0625 = .2500$$

$$F(.2) = 6.25(.2)^2 - 1.25(.2) + .0625 = .0625$$

Thus

$$\begin{aligned} P[.2 \le X \le .3] &= F(.3) - F(.2) \\ &= .2500 - .0625 = .1875 \end{aligned}$$

Note that this agrees with the result obtained in Example 4.1.1 using direct integration. Note also that $F(.3)$ gives the area to the left of .3 shown in Fig. 4.3(a); $F(.2)$ gives the area to the left of .2 shown in Fig. 4.3(b). When we form the difference $F(.3) - F(.2)$, we naturally obtain the area between .2 and .3 given in Fig. 4.3(c).

Uniform Distribution

Perhaps the simplest continuous distribution with which to work is the *uniform* distribution. This distribution parallels the discrete uniform distribution presented in Exercise 32 of Chap. 3 in that, in a sense, events occur with equal or uniform probability. Since it is easy and instructive to develop the properties of this family of random variables directly from the definition, we leave the derivations to you. Important properties and applications are given in Exercises 5, 6, 10, 11, 18, 19, and 23.

4.2 EXPECTATION AND DISTRIBUTION PARAMETERS

In this section we define the term *expected value for continuous random variables*. We also discuss how to use the definition to find the moment generating function, the mean and the variance of a variable of the continuous type. As you will see, the definition parallels that given in the discrete case with the summation operation being replaced by integration.

Definition 4.2.1 (Expected value). Let X be a continuous random variable with density f. Let $H(X)$ be a random variable. The expected value of $H(X)$, denoted by $E[H(X)]$, is given by

$$E[H(X)] = \int_{-\infty}^{\infty} H(x)f(x)\,dx$$

provided

$$\int_{-\infty}^{\infty} |H(x)|f(x)\,dx$$

is finite.

We illustrate the use of this definition by finding the mean and variance of the random variable X of Example 4.1.1.

Example 4.2.1. The density for X, the lead concentration in gasoline in grams per liter is given by

$$f(x) = 12.5x - 1.25 \qquad .1 \le x \le .5$$

The mean or expected value of X is

$$\begin{aligned}
\mu = E[X] &= \int_{-\infty}^{\infty} xf(x)\,dx \\
&= \int_{.1}^{.5} x(12.5x - 1.25)\,dx \\
&= \frac{12.5x^3}{3} - \frac{1.25x^2}{2}\Bigg|_{.1}^{.5} \\
&= \left[\frac{(12.5)(.5)^3}{3} - \frac{1.25(.5)^2}{2}\right] - \left[\frac{12.5(.1)^3}{3} - \frac{1.25(.1)^2}{2}\right] \\
&\doteq .3667 \text{ g/liter}
\end{aligned}$$

Since integration is over an interval of finite length

$$\int_{-\infty}^{\infty} |x| f(x)\, dx$$

exists. We can conclude that on the average, a liter of gasoline contains approximately .3667 g of lead. How much variability is there from liter to liter? To answer this question, we find $E[X^2]$ and apply Theorem 3.3.2 to find the variance of X.

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)\, dx \\ &= \int_{.1}^{.5} x^2(12.5x - 1.25)\, dx \\ &= \frac{12.5x^4}{4} - \frac{1.25x^3}{3}\bigg|_{.1}^{.5} \doteq .1433 \end{aligned}$$

By Theorem 3.3.2

$$\operatorname{Var} X = E[X^2] - (E[X])^2 \doteq .1433 - (.3667)^2 \doteq .00883$$

The standard deviation of X is

$$\sigma = \sqrt{\operatorname{Var} X} = \sqrt{.00883} = .09396 \text{ g/liter}$$

As in the discrete case, the moment generating function for a continuous random variable X is defined as $E[e^{tX}]$ provided this expectation exists for t in some open interval about 0. Its use is illustrated in the following example.

Example 4.2.2. The spontaneous flipping of a bit stored in a computer memory is called a "soft fail." Let X denote the time in millions of hours before the first soft fail is observed. Suppose that the density for X is given by

$$f(x) = e^{-x} \qquad x > 0$$

The mean and variance for X can be found directly using the method of Example 4.2.1. However, to find $E[X]$ and $E[X^2]$ integration by parts is required. This method of integration, while not difficult, is time-consuming. Let us find the moment generating function for X and use it to compute the mean and variance. By definition

$$m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

In this case

$$\begin{aligned} m_X(t) &= \int_0^{\infty} e^{tx} e^{-x}\, dx \\ &= \int_0^{\infty} e^{(t-1)x}\, dx \\ &= \frac{1}{t-1} e^{(t-1)x}\bigg|_0^{\infty} \end{aligned}$$

Assume that $|t| < 1$. This guarantees that the exponent $(t - 1)x < 0$ allowing us to evaluate the above integral. In particular

$$m_X(t) = \frac{1}{1-t} \qquad |t| < 1$$

Since $e^{tx} > 0$, $|e^{tx}| = e^{tx}$. Thus the above argument has shown that

$$\int_{-\infty}^{\infty} |e^{tx}| f(x)\, dx$$

exists, as required in Definition 4.2.1. To use $m_X(t)$ to find $E[X]$ and $E[X^2]$, we apply Theorem 3.4.2. Note that

$$\frac{dm_X(t)}{dt} = \frac{d(1-t)^{-1}}{dt} = (1-t)^{-2}$$

$$\frac{d^2m_X(t)}{dt^2} = 2(1-t)^{-3}$$

$$E[X] = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = 1$$

$$E[X^2] = \left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = 2$$

$$\operatorname{Var} X = E[X^2] - (E[X])^2 = 2 - 1^2 = 1$$

The average or mean time that one must wait to observe the first soft fail is one million hours. The variance in waiting time is 1 and the standard deviation is one million hours.

To find the distribution parameters μ, σ^2, and σ, we can use either Definition 4.2.1 or the moment generating function technique. In practice, use whichever method is easier. Exercises 20 and 21 point out some interesting aspects of the mean and variance in the case of a continuous random variable.

4.3 GAMMA DISTRIBUTION

In this section we consider the gamma distribution. This distribution is especially important in that it allows us to define two families of random variables, the exponential and chi-squared, that are used extensively in applied statistics. The theoretical basis for the gamma distribution is the gamma function, a mathematical function defined in terms of an integral.

Definition 4.3.1 (Gamma function). The function Γ defined by

$$\Gamma(\alpha) = \int_0^{\infty} z^{\alpha-1} e^{-z}\, dz \qquad \alpha > 0$$

is called the gamma function.

Theorem 4.3.1 presents two numerical properties of the gamma function that are useful in evaluating the function for various values of α. Its proof is outlined in Exercise 26.

Theorem 4.3.1 (Properties of the gamma function)

1. $\Gamma(1) = 1$.
2. For $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

The use of Theorem 4.3.1 is illustrated in the next example.

Example 4.3.1

(*a*) Evaluate $\int_0^\infty z^3 e^{-z}\,dz$. To evaluate this integral using the methods of elementary calculus requires repeated applications of integration by parts. To evaluate the integral quickly, rewrite it as

$$\int_0^\infty z^3 e^{-z}\,dz = \int_0^\infty z^{4-1} e^{-z}\,dz$$

The integral on the right is $\Gamma(4)$. By applying Theorem 4.3.1 repeatedly it can be seen that

$$\begin{aligned}\int_0^\infty z^3 e^{-z}\,dz = \Gamma(4) &= 3 \cdot \Gamma(3)\\ &= 3 \cdot 2 \cdot \Gamma(2)\\ &= 3 \cdot 2 \cdot 1 \cdot \Gamma(1)\\ &= 3 \cdot 2 \cdot 1 = 6\end{aligned}$$

(*b*) Evaluate $\int_0^\infty (1/54)x^2 e^{-x/3}\,dx$. To evaluate this integral, we make a change of variable, a technique that is used extensively in deriving the properties of the gamma distribution. In particular, let $z = x/3$ or $3z = x$. Then $3\,dz = dx$ and the problem becomes

$$\begin{aligned}\int_0^\infty (1/54)\,x^2 e^{-x/3}\,dx &= \int_0^\infty 1/54(3z)^2 e^{-z} 3\,dz\\ &= 27/54 \int_0^\infty z^2 e^{-z}\,dz\end{aligned}$$

However,

$$\begin{aligned}\int_0^\infty z^2 e^{-z}\,dz = \int_0^\infty z^{3-1} e^{-z}\,dz &= \Gamma(3)\\ &= 2 \cdot \Gamma(2)\\ &= 2 \cdot 1 \cdot \Gamma(1)\\ &= 2 \cdot 1 = 2\end{aligned}$$

Thus

$$\int_0^\infty (1/54)\,x^2 e^{-x/3}\,dx = 27/54 \cdot 2 = 1$$

Note that since the nonnegative function

$$f(x) = (1/54)x^2e^{-x/3}$$

has been shown to integrate to 1, it can be thought of as being a density for a continuous random variable X.

It is now possible to define the gamma distribution.

Definition 4.3.2 (Gamma distribution). A random variable X with density

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta} \qquad \begin{array}{l} x > 0 \\ \alpha > 0 \\ \beta > 0 \end{array}$$

is said to have a gamma distribution with parameters α and β.

Although the mean and variance of a gamma random variable can be found easily from the definitions of these parameters (see Exercise 31), we shall use the moment generating function technique. As you will see later, it is very helpful to know the form of the moment generating function for a random variable.

Theorem 4.3.2. Let X be a gamma random variable with parameters α and β. Then

1. The moment generating function for X is given by

$$m_X(t) = (1 - \beta t)^{-\alpha} \qquad t < 1/\beta$$

2. $E[X] = \alpha\beta$
3. $\operatorname{Var} X = \alpha\beta^2$

Proof. By definition

$$m_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}\,dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-(1/\beta - t)x}\,dx$$

Let $z = (1 - \beta t)x/\beta$ so that $x = \beta z/(1 - \beta t)$. Then $\beta\,dz = (1 - \beta t)\,dx$ and $dx = \beta\,dz/(1 - \beta t)$. By making these substitutions, we obtain

$$m_X(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty \left(\frac{\beta z}{1 - \beta t}\right)^{\alpha-1} e^{-z}\frac{\beta\,dz}{(1 - \beta t)}$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\frac{\beta^\alpha}{(1 - \beta t)^\alpha}\int_0^\infty z^{\alpha-1}e^{-z}\,dz$$

The integral on the right is $\Gamma(\alpha)$. Therefore

$$m_X(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\frac{\beta^\alpha}{(1 - \beta t)^\alpha}\Gamma(\alpha) = (1 - \beta t)^{-\alpha}$$

We restrict t to be less than $1/\beta$ to avoid possible division by 0.

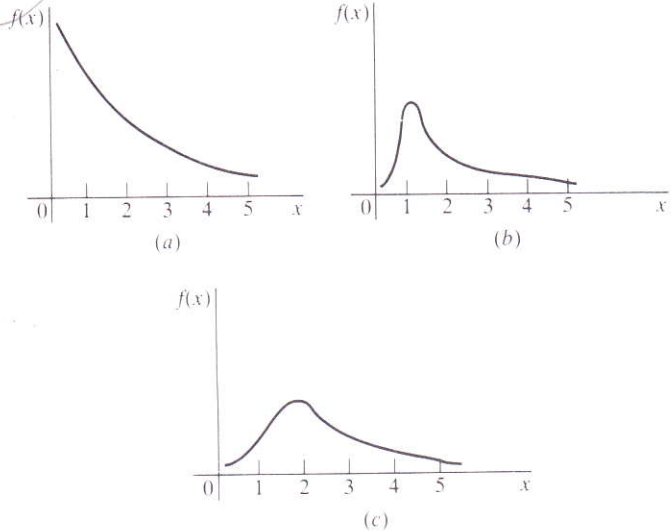


FIGURE 4.4
(*a*) $\alpha = 1$, $\beta = 1$, $\mu_X = 1$, $\sigma_X^2 = 1$. (*b*) $\alpha = 2$, $\beta = 1$, $\mu_X = 2$, $\sigma_X^2 = 2$. (*c*) $\alpha = 2$, $\beta = 2$, $\mu_X = 4$, $\sigma_X^2 = 8$.

The proofs of parts 2 and 3 are straightforward applications of Theorem 3.4.2 and are left as exercises. (See Exercise 30.)

Figure 4.4 shows the graphs of some gamma densities for a few values of α and β. Note that α and β both play a role in determining the mean and the variance of the random variable. Note also that the curves are not symmetric and are located entirely to the right of the vertical axis. It can be shown that for $\alpha > 1$, the maximum value of the density occurs at the point $x = (\alpha - 1)\beta$. (See Exercise 32.)

Exponential Distribution

As mentioned earlier, the gamma distribution gives rise to a family of random variables known as the *exponential* family. These variables are each gamma random variables with $\alpha = 1$. The density for an exponential random variable therefore assumes the form

Exponential density

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \qquad x > 0$$

$$\beta > 0$$

The graph of a typical exponential density is shown in Fig. 4.4(a). This distribution arises often in practice in conjunction with the study of Poisson processes which were discussed in Sec. 3.8. Recall that in a Poisson process discrete events are being observed over a continuous time interval. If we let W denote the time of the occurrence of the first event, then W is a continuous random variable. Theorem 4.3.3 shows that W has an exponential distribution.

Theorem 4.3.3. Consider a Poisson process with parameter λ. Let W denote the time of the occurrence of the first event. W has an exponential distribution with $\beta = 1/\lambda$.

Proof. The distribution function F for W is given by

$$F(w) = P[W \leq w] = 1 - P[W > w]$$

The first occurrence of the event will take place after time w only if no occurrences of the event are recorded in the time interval $[0, w]$. Let X denote the number of occurrences of the event in this time interval. X is a Poisson random variable with parameter λw. Thus

$$P[W > w] = P[X = 0] = \frac{e^{-\lambda w}(\lambda w)^0}{0!} = e^{-\lambda w}$$

By substitution, we obtain

$$F(w) = 1 - P[W > w] = 1 - e^{-\lambda w}$$

Since, in the continuous case, the derivative of the cumulative distribution function is the density (see Exercise 14),

$$F'(w) = f(w) = \lambda e^{-\lambda w}$$

This is the density for an exponential random variable with $\beta = 1/\lambda$.

The next example illustrates the use of this theorem.

Example 4.3.2. Some strains of paramecia produce and secrete "killer" particles that will cause the death of a sensitive individual if contact is made. All paramecia unable to produce killer particles are sensitive. The mean number of killer particles emitted by a killer paramecium is 1 every five hours. In observing such a paramecium what is the probability that we must wait at most four hours before the first particle is emitted? Considering the measurement unit to be one hour, we are observing a Poisson process with $\lambda = 1/5$. By Theorem 4.3.3 W, the time at which the first killer particle is emitted, has an exponential distribution with $\beta = 1/\lambda = 5$. The density for W is

$$f(w) = (1/5)e^{-w/5} \qquad w > 0$$

The desired probability is given by

$$\begin{aligned} P[W \leq 4] &= \int_0^4 (1/5)e^{-w/5}\,dw \\ &= -e^{-w/5}\Big|_0^4 \\ &= 1 - e^{-4/5} \doteq .5507 \end{aligned}$$

Since an exponential random variable is also a gamma random variable, the average time that we must wait until the first killer particle is emitted is

$$E[W] = \alpha\beta = 1 \cdot 5 = 5 \text{ hours}$$

Chi-Squared Distribution

The gamma distribution gives rise to another important family of random variables, namely, the *chi-squared* family. This distribution is used extensively in applied statistics. Among other things, it provides the basis for making inferences about the variance of a population based on a sample. At this time we consider only the theoretical properties of the chi-squared distribution. You will see many examples of its use in later chapters.

Definition 4.3.3 (Chi-squared distribution). Let X be a gamma random variable with $\beta = 2$ and $\alpha = \gamma/2$ for γ a positive integer. X is said to have a chi-squared distribution with γ degrees of freedom. We denote this variable by X^2_γ.

Note that a chi-squared random variable is completely specified by stating its degrees of freedom. By applying Theorem 4.3.2 it is seen that the mean of a chi-squared random variable is γ, its degrees of freedom; its variance is 2γ, twice its degrees of freedom. Figure 4.4(c) gives the graph of the density of a chi-squared random variable with four degrees of freedom.

Since the chi-squared distribution arises so often in practice, extensive tables of its cumulative distribution function have been derived. One such table is Table IV of App. A. In the table, degrees of freedom appear as row headings, probabilities appear as column headings, and points associated with those probabilities are listed in the body of the table. Notationally we shall use χ^2_r to denote that point associated with a chi-squared random variable such that

$$P\left[X^2_\gamma \geq \chi^2_r\right] = r$$

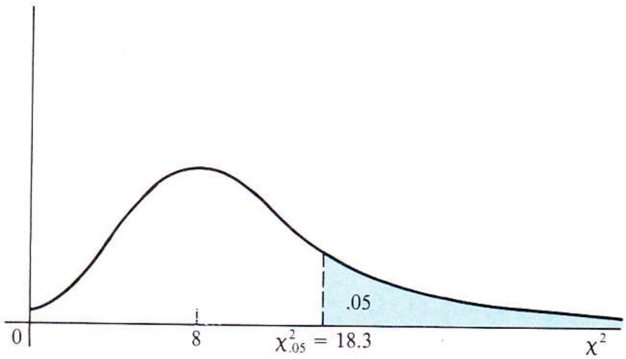


FIGURE 4.5
$P[X^2_{10} \geq \chi^2_{.05}] = .05$ and $P[X^2_{10} < \chi^2_{.05}] = .95$.

That is, χ^2_r is the point such that the area to its *right* is r. Technically speaking, we should write $\chi^2_{r,\gamma}$ since the value of the point does depend on both the probability desired and the number of degrees of freedom associated with the random variable. However, in applications, the value of γ will be obvious. Therefore, to simplify notation we use only a single subscript. The use of this notation is illustrated in the following example.

Example 4.3.3. Consider a chi-squared random variable with 10 degrees of freedom. Find the value of $\chi^2_{.05}$. This point is shown in Fig. 4.5. By definition, the area to the right of this point is .05; the area to its left is .95. The column probabilities in Table IV give the area to the *left* of the point listed. Thus, to find $\chi^2_{.05}$, we look in row 10 and column .95 and see that $\chi^2_{.05} = 18.3$.

4.4 NORMAL DISTRIBUTION

The normal distribution is a distribution that underlies many of the statistical methods used in data analysis. It was first described in 1733 by De Moivre as being the limiting form of the binomial density as the number of trials becomes infinite. This discovery did not get much attention, and the distribution was "discovered" again by both Laplace and Gauss a half-century later. Both men dealt with problems of astronomy, and each derived the normal distribution as a distribution that seemingly described the behavior of errors in astronomical measurements. The distribution is often referred to as the "gaussian" distribution.

Definition 4.4.1 (Normal distribution). A random variable X with density

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2} \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \\ \sigma > 0 \end{array}$$

is said to have a normal distribution with parameters μ and σ.

One implication of this definition is that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2}\, dx = 1$$

To verify this requires a transformation to polar coordinates. This technique is beyond the mathematical level assumed here. A detailed proof can be found in [47]. Note that Definition 4.4.1 states only that μ is a real number and that σ is positive. As you might suspect from the notation used, the parameters that appear in the equation for the density for a normal random variable are, in fact, its mean and its standard deviation. This can be verified once we know the moment generating function for X. Our next theorem gives us the form for this important function.

Theorem 4.4.1. Let X be normally distributed with parameters μ and σ. The moment generating function for X is given by

$$m_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Proof. By definition

$$\begin{aligned} m_X(t) = E[e^{tX}] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2}\,dx \\ &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{tx-(1/2)[(x-\mu)/\sigma]^2}\,dx \end{aligned}$$

We complete the square in the exponent as follows:

$$\begin{aligned} &tx - (1/2)[(x-\mu)/\sigma]^2 \\ &\quad = tx - [1/2\sigma^2](x^2 - 2\mu x + \mu^2) \\ &\quad = -[1/2\sigma^2](x^2 - 2\mu x - 2\sigma^2 tx + \mu^2) \\ &\quad = -[1/2\sigma^2]\left[x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2 + \mu^2\right] \\ &\quad = -[1/2\sigma^2][x - (\mu + \sigma^2 t)]^2 + [1/2\sigma^2](\mu^2 + 2\mu\sigma^2 t + \sigma^4 t^2) - [1/2\sigma^2]\mu^2 \\ &\quad = -[1/2\sigma^2][x - (\mu + \sigma^2 t)]^2 + \mu t + \sigma^2 t^2/2 \end{aligned}$$

By substituting this expression into that for $m_X(t)$, we obtain

$$\begin{aligned} m_X(t) &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-[1/2\sigma^2][x-(\mu+\sigma^2 t)]^2 + \mu t + \sigma^2 t^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-(1/2)[x-(\mu+\sigma^2 t)/\sigma]^2} e^{\mu t + \sigma^2 t^2/2}\,dx \\ &= e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[x-(\mu+\sigma^2 t)/\sigma]^2}\,dx \end{aligned}$$

The function

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[x-(\mu+\sigma^2 t)/\sigma]^2}$$

is the density for a normal random variable with parameters σ and $\mu + \sigma^2 t$ and thus integrates to 1. This implies that

$$m_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

as claimed.

It is now easy to show that the parameters that appear in the definition of the normal density are actually the mean and the standard deviation of the variable.

Theorem 4.4.2. Let X be a normal random variable with parameters μ and σ. Then μ is the mean of X and σ is its standard deviation.

Proof. The moment generating function for X is

$$m_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

and

$$\frac{dm_X(t)}{dt} = e^{\mu t + \sigma^2 t^2/2}(\mu + \sigma^2 t)$$

By Theorem 3.4.2, the mean of X is given by

$$E[X] = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = e^{\mu \cdot 0 + \sigma^2 0^2/2}(\mu + \sigma^2 \cdot 0) = \mu$$

as claimed. The proof of the remainder of the theorem is left as an exercise (Exercise 44).

The graph of the density of a normal random variable is a symmetric, bell-shaped curve centered at its mean. The points of inflection occur at $\mu \pm \sigma$. These facts can be verified easily (see Exercise 43).

Example 4.4.1. One of the major contributors to air pollution is hydrocarbons emitted from the exhaust system of automobiles. Let X denote the number of grams of hydrocarbons emitted by an automobile per mile. Assume that X is normally distributed with a mean of 1 gram and a standard deviation of .25 gram. The density for X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}(.25)} e^{-(1/2)[(x-1)/.25]^2}$$

The graph of this density is a symmetric, bell-shaped curve centered at $\mu = 1$ with inflection points at $\mu \pm \sigma$, or $1 \pm .25$. A sketch of the density is given in Fig. 4.6.

One point must be made. Theoretically speaking, a normal random variable must be able to assume any value whatsoever. This is clearly unrealistic here. It is impossible for an automobile to emit a negative amount of hydrocarbons. When we

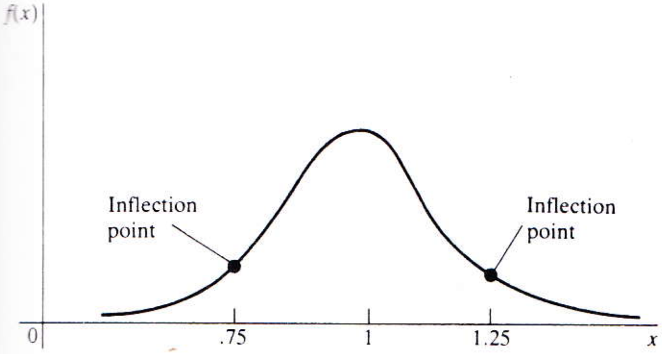


FIGURE 4.6
Graph of the density for a normal random variable with mean 1 and standard deviation .25.

say that X is normally distributed, we mean that over the range of physically reasonable values of X, the given normal curve yields acceptable probabilities. With this understanding we can at least approximate, for example, the probability that a randomly selected automobile will emit between .9 and 1.54 grams of hydrocarbons by finding the area under the graph of f between these two points.

There are infinitely many normal random variables each uniquely characterized by the two parameters μ and σ. To calculate probabilities associated with a specific normal curve requires that one integrate the normal density over a particular interval. However, the normal density is not integrable in closed form. To find areas under the normal curve requires the use of numerical integration techniques. A simple algebraic transformation is employed to overcome this problem. By means of this transformation, called the *standardization procedure*, any question about any normal random variable can be transformed to an equivalent question concerning a normal random variable with mean 0 and standard deviation 1. This particular normal random variable is denoted by Z and is called the *standard normal* variable.

Theorem 4.4.3 (Standardization theorem). Let X be normal with mean μ and standard deviation σ. The variable $(X - \mu)/\sigma$ is standard normal.

You have already verified that the transformation yields a random variable with mean 0 and standard deviation 1 (see Chap. 3, Exercise 21). To prove that the transformed variable is normal requires the use of moment generating function techniques to be introduced in Chap. 7.

The cumulative distribution function for the standard normal random variable is given in Table V of App. A. The use of the standardization theorem and this table is illustrated in the following example.

Example 4.4.2. Let X denote the number of grams of hydrocarbons emitted by an automobile per mile. Assuming that X is normal with $\mu = 1$ gram and $\sigma = .25$

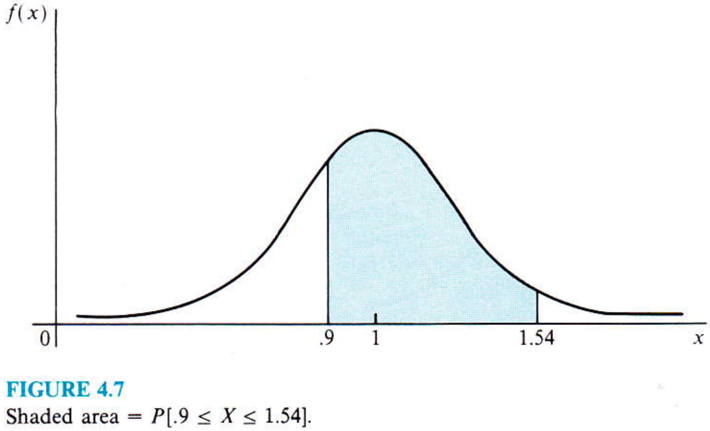


FIGURE 4.7
Shaded area $= P[.9 \le X \le 1.54]$.

gram, find the probability that a randomly selected automobile will emit between .9 and 1.54 grams of hydrocarbons per mile. The desired probability is shown in Fig. 4.7. To find $P[.9 \le X \le 1.54]$, we first standardize by subtracting the mean of 1 and dividing by the standard deviation of .25 across the inequality. That is,

$$P[.9 \le X \le 1.5] = P[(.9 - 1)/.25 \le (X - 1)/.25 \le (1.54 - 1)/.25]$$

The random variable $(X - 1)/.25$ is now Z. Therefore the problem is to find $P[-.4 \le Z \le 2.16]$ from Table V. We first express the desired probability in terms of the cumulative distribution as follows:

$$\begin{aligned} P[-.4 \le Z \le 2.16] &= P[Z \le 2.16] - P[Z < -.4] \\ &= P[Z \le 2.16] - P[Z \le -.4] \qquad (Z \text{ is continuous}) \\ &= F(2.16) - F(-.4) \end{aligned}$$

$F(2.16)$ is found by locating the first two digits (2.1) in the column headed z; since the third digit is 6, the desired probability of .9846 is found in the row labeled 2.1 and the column labeled .06. Similarly, $F(-.4)$ or .3446 is found in the row labeled -0.4 and the column labeled .00. We now see that the probability that a randomly selected automobile will emit between .9 and 1.54 grams of hydrocarbons per mile is

$$\begin{aligned} P[.9 \le X \le 1.54] &= P[-.4 \le Z \le 2.16] \\ &= F(2.16) - F(-.4) \\ &= .9846 - .3446 = .64 \end{aligned}$$

Interpreting this probability as a percentage, we can say that 64% of the automobiles in operation emit between .9 and 1.54 grams of hydrocarbons per mile driven.

We shall have occasion to read Table V in reverse. That is, given a particular probability r we shall need to find the point with r of the area to its right. This point is denoted by z_r. Thus, notationally, z_r denotes that point associated with a standard normal random variable such that

$$P[Z \ge z_r] = r$$

To see how this need arises, consider Example 4.4.3.

Example 4.4.3. Let X denote the amount of radiation that can be absorbed by an individual before death ensues. Assume that X is normal with a mean of 500 roentgens and a standard deviation of 150 roentgens. Above what dosage level will only 5% of those exposed survive? Here we are asked to find the point x_0 shown in Fig. 4.8. In terms of probabilities, we want to find the point x_0 such that

$$P[X \ge x_0] = .05$$

Standardizing gives

$$\begin{aligned} P[X \ge x_0] &= P\left[\frac{X - 500}{150} \ge \frac{x_0 - 500}{150}\right] \\ &= P\left[Z \ge \frac{x_0 - 500}{150}\right] = .05 \end{aligned}$$

Thus $(x_0 - 500)/150$ is the point on the standard normal curve with 5% of the area under the curve to its right and 95% to its left. That is, $(x_0 - 500)/150$ is the point $z_{.05}$. From Table V, the numerical value of this point is approximately 1.645 (we

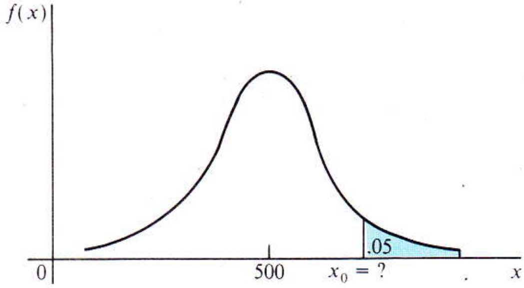


FIGURE 4.8
$P[X \geq x_0] = .05.$

have interpolated). Equating these, we get

$$\frac{x_0 - 500}{150} = 1.645$$

Solving this equation for x_0 gives the desired dosage level:

$$x_0 = 150(1.645) + 500 = 746.75 \text{ roentgens}$$

4.5 NORMAL PROBABILITY RULE AND CHEBYSHEV'S INEQUALITY (OPTIONAL)

It is sometimes useful to have a quick way of determining which values of a random variable are common and which are considered to be rare. In the case of a normally distributed random variable, a rule of thumb, called the *normal probability rule*, can be developed easily. This rule is given in Theorem 4.5.1.

Theorem 4.5.1 (Normal probability rule). Let X be normally distributed with parameters μ and σ. Then

$$P[-\sigma < X - \mu < \sigma] \doteq .68$$
$$P[-2\sigma < X - \mu < 2\sigma] \doteq .95$$
$$P[-3\sigma < X - \mu < 3\sigma] \doteq .99$$

Proof. Note that division by σ yields

$$P[-\sigma < X - \mu < \sigma] = P\left[-1 < \frac{X - \mu}{\sigma} < 1\right]$$

By Theorem 4.4.3, $(X - \mu)/\sigma$ follows the standard normal distribution. From Table V of App. A,

$$P[-1 < Z < 1] = .8413 - .1587 = .6826$$

This probability can be rounded to .68. The other results given in the theorem are proved similarly.

The normal probability rule can be expressed in terms of percentages. In particular, it implies that in repeated sampling from a normal distribution approximately 68% of the observed values of X should lie within 1 standard deviation of its mean; 95% should lie within 2 standard deviations, and 99% within 3 standard deviations of the mean. Thus an observed value that falls farther than 3 standard deviations from μ is indeed rare since such values occur with probability .01. This rule will be used later to obtain a quick estimate of the standard deviation of a normally distributed random variable.

Chebyshev's Inequality

A second rule of thumb that can be used to gauge the rarity of observed values of a random variable is *Chebyshev's inequality*. This inequality was derived by the Russian probabilist P. L. Chebyshev (Tchebysheff, 1821–1894). The inequality differs from the normal probability rule in that it does *not* require that the random variable involved be normally distributed. Although we shall prove the theorem in the continuous setting, continuity is *not* required. The inequality holds for any random variable.

Theorem 4.5.2 (Chebyshev's inequality). Let X be a random variable with mean μ and standard deviation σ. Then for any positive number k,

$$P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2}$$

Proof. Assume that X is continuous with mean μ, standard deviation σ, and density f. By definition,

$$\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

Let $k > 0$, $c = k^2\sigma^2$, and note that

$$\sigma^2 = \int_{-\infty}^{\mu - \sqrt{c}} (x - \mu)^2 f(x)\, dx + \int_{\mu - \sqrt{c}}^{\mu + \sqrt{c}} (x - \mu)^2 f(x)\, dx$$
$$+ \int_{\mu + \sqrt{c}}^{\infty} (x - \mu)^2 f(x)\, dx$$

Since $(x - \mu)^2 f(x) \geq 0$,

$$\int_{\mu - \sqrt{c}}^{\mu + \sqrt{c}} (x - \mu)^2 f(x)\, dx \geq 0$$

and thus

$$\sigma^2 \geq \int_{-\infty}^{\mu - \sqrt{c}} (x - \mu)^2 f(x)\, dx + \int_{\mu + \sqrt{c}}^{\infty} (x - \mu)^2 f(x)\, dx$$

Note that over both regions of integration, $(x - \mu)^2 \geq c$ and so it can be concluded that

$$\sigma^2 \geq \int_{-\infty}^{\mu - \sqrt{c}} cf(x)\, dx + \int_{\mu + \sqrt{c}}^{\infty} cf(x)\, dx$$

In terms of probabilities,

$$\sigma^2 \geq cP\left[X \leq \mu - \sqrt{c}\right] + cP\left[X \geq \mu + \sqrt{c}\right]$$

or

$$\sigma^2 \geq c\left\{P\left[X - \mu \leq -\sqrt{c}\right] + P\left[X - \mu \geq \sqrt{c}\right]\right\}$$

This inequality can be rewritten to conclude that

$$P\left[X - \mu \leq -\sqrt{c}\right] + P\left[X - \mu \geq \sqrt{c}\right] \leq \sigma^2/c$$

or that

$$P\left[-\sqrt{c} \leq X - \mu \leq \sqrt{c}\right] \geq 1 - \sigma^2/c$$

Since $c = k^2\sigma^2$ where k and c are each nonnegative, $c = k\sigma$. Substitution yields

$$P[-k\sigma \leq X - \mu \leq k\sigma] \geq 1 - \frac{1}{k^2}$$

or

$$P[|X - \mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}$$

Since X is continuous, we can conclude that

$$P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2}$$

as claimed. The proof in the discrete case is similar with summation replacing integration.

Some examples will clarify the difference between Theorems 4.5.1 and 4.5.2.

Example 4.5.1. The viscosity of a fluid can be measured roughly by dropping a small ball into a calibrated tube containing the fluid and observing X, the time that it takes for the ball to drop a measured distance. Assume that this random variable is normally distributed with a mean of 20 s and a standard deviation of .5 s. By the normal probability rule, approximately 95% of the observed values of X will lie within 1 s (2 standard deviations) of the mean. That is, X will fall between 19 and 21 s with probability .95. Since Chebyshev's inequality applies to any random variable, it is appropriate here. This inequality guarantees that X will fall between 19 and 21 s (within $k = 2$ standard deviations of its mean) with probability *at least* $1 - 1/k^2 = .75$. Note that when the random variable in question is normally distributed, the normal probability rule yields a stronger statement than does Chebyshev's inequality.

Example 4.5.2. The safety record of an industrial plant is measured in terms of M, the total man-hours worked without a serious accident. Past experience indicates that M has a mean of 2 million with a standard deviation of .1 million. A serious accident has just occurred. Would it be unusual for the next serious accident to occur within the next 1.6 million man-hours? To answer this question, we must assess $P[M \leq 1.6]$. Since we have no reason to assume that M is normally distributed, the normal probability rule is inappropriate here. However, we know from Chebyshev's inequality with $k = 4$ that

$$P[1.6 < M < 2.4] \geq 1 - \tfrac{1}{16} = .9375$$

This implies that

$$P[M \leq 1.6] + P[M \geq 2.4] \leq .0625$$

Since it is possible for M to exceed 2.4, we can safely say that

$$P[M \leq 1.6] < .0625$$

No stronger statement can be made without some knowledge of the shape of the density of M. However, if it is known that the density is symmetric, then we can go one step further and state that

$$P[M \leq 1.6] \leq .0625/2 = .03125$$

4.6 NORMAL APPROXIMATIONS

We saw in Sec. 3.8 that the Poisson density can be used to approximate binomial probabilities when n is large. The normal curve can be used for the same purpose. To see how this can be done, we consider four binomial random variables each with probability of success .4 but with differing values for n. The densities for these variables, obtained from Table I of App. A, together with a sketch for each, are given in Fig. 4.9(a) to (d).

The point to note from these diagrams is made in Fig. 4.9(d). Namely, it is not hard to imagine a smooth bell curve that closely fits the block diagram shown. This suggests that binomial probabilities represented by one or more blocks in the diagram can be approximated reasonably well by a carefully selected area under an appropriately chosen normal curve. Which of the infinitely many normal curves is appropriate? Common sense indicates that the normal variable selected should have the same mean and variance as the binomial variable that it approximates. Theorem 4.6.1 summarizes these ideas.

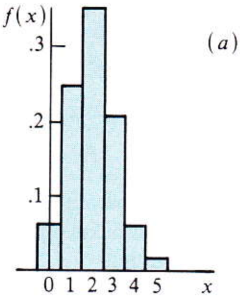
Theorem 4.6.1 (Normal approximation to the binomial distribution). Let X be binomial with parameters n and p. For large n, X is approximately normal with mean np and variance $np(1-p)$.

The proof of this theorem is based on the Central Limit Theorem which will be considered in Chap. 7. Admittedly, Theorem 4.5.1 is a bit vague in the sense that the word *large* is not well defined. In the strictest mathematical sense, large means as n approaches infinity. For most practical purposes the approximation is acceptable for values of n and p such that either $p \leq .5$ and $np > 5$ or $p > .5$ and $n(1-p) > 5$.

Example 4.6.1. A study is performed to investigate the connection between maternal smoking during pregnancy and birth defects in children. Of the mothers studied, 40% smoke and 60% do not. When the babies were born, 20 were found to have some sort of birth defect. Let X denote the number of children whose mother smoked while pregnant. If there is no relationship between maternal smoking and birth defects, then X is binomial with $n = 20$ and $p = .4$. What is the probability that 12 or more of the affected children had mothers who smoked?

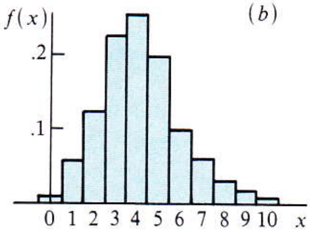
$n = 5$
$p = .4$

x	$f(x)$
0	.0778
1	.2592
2	.3456
3	.2304
4	.0768
5	.0102



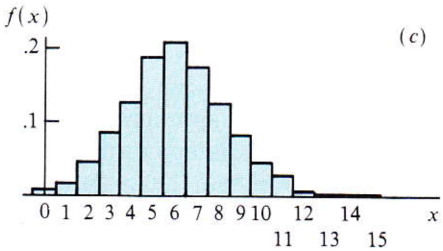
$n = 10$
$p = .4$

x	$f(x)$
0	.0060
1	.0404
2	.1209
3	.2150
4	.2508
5	.2007
6	.1114
7	.0425
8	.0106
9	.0016
10	.0001



$n = 15$
$p = .4$

x	$f(x)$	x	$f(x)$
0	.0005	8	.1181
1	.0047	9	.0612
2	.0219	10	.0245
3	.0634	11	.0074
4	.1268	12	.0016
5	.1859	13	.0003
6	.2066	14	~ 0
7	.1771	15	~ 0



$n = 20$
$p = .4$

x	$f(x)$	x	$f(x)$
0	~ 0	11	.0710
1	.0005	12	.0355
2	.0031	13	.0145
3	.0124	14	.0049
4	.0350	15	.0013
5	.0746	16	.0003
6	.1244	17	~ 0
7	.1659	18	~ 0
8	.1797	19	~ 0
9	.1597	20	~ 0
10	.1172		

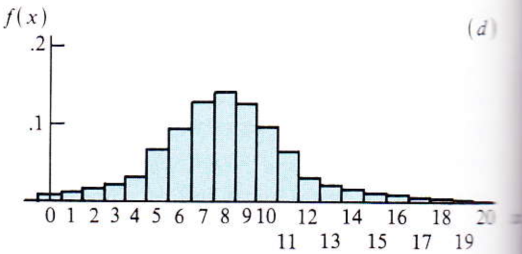


FIGURE 4.9
Density for X binomial: (a) $n = 5$, $p = .4$; (b) $n = 10$, $p = .4$; (c) $n = 15$, $p = .4$; (d) $n = 20$, $p = .4$.

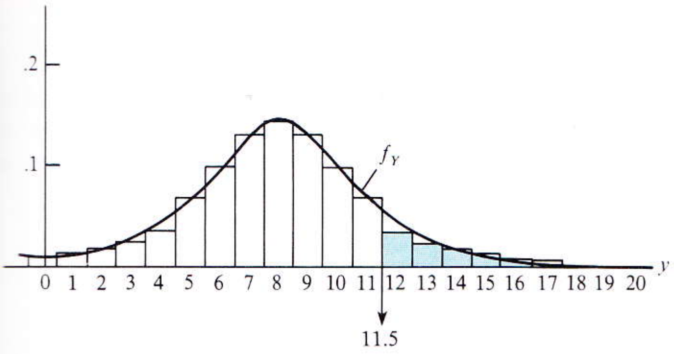


FIGURE 4.10
$P[X \geq 12]$ = area of shaded blocks $\doteq$ area under curve beyond 11.5.

To answer this question, we need to find $P[X \geq 12]$ under the assumption that X is binomial with $n = 20$ and $p = .4$. This probability, .0565, can be found from Table I of App. A. Note that since $p = .4 \leq .5$ and $np = 20(.4) = 8 > 5$, the normal approximation should give a result quite close to .0565. We shall approximate probabilities associated with X using a normal random variable Y with mean $np = 20(.4) = 8$ and standard deviation $\sqrt{np(1-p)} = \sqrt{20(.4)(.6)} = \sqrt{4.8}$.

The exact probability of .0565 is given by the sum of the areas of the blocks centered at 12, 13, 14, 15, 16, 17, 18, 19, and 20 as shown in Fig. 4.10. The approximate probability is given by the area under the normal curve shown above 11.5. That is,

$$P[X \geq 12] \doteq P[Y \geq 11.5]$$

The number .5 is called the *half-unit correction* for continuity. It is subtracted from 12 in the approximation because otherwise half of the area of the block centered at 12 will be inadvertently ignored, leading to an unnecessary error in the calculation. From this point on, the calculation is routine.

$$\begin{aligned} P[X \geq 12] &\doteq P[Y \geq 11.5] \\ &= P\left[\frac{Y-8}{\sqrt{4.8}} \geq \frac{11.5-8}{\sqrt{4.8}}\right] \\ &= P[Z \geq 1.59] \\ &= 1 - .9441 = .0559 \end{aligned}$$

Note that even with n as small as 20, the approximated value of .0559 compares quite favorably with the exact value of .0565. In practice, of course, one would not approximate a probability that could be found directly from a binomial table. This was done here only for comparative purposes.

Poisson probabilities can also be approximated with the help of a normal curve. The method for doing so is outlined in Exercise 57.

4.7 WEIBULL DISTRIBUTION AND RELIABILITY

In 1951, W. Weibull introduced a distribution that has been found to be useful in a variety of physical applications. It arises quite naturally in the study of reliability as we shall show. The most general form for the Weibull density is given by

$$f(x) = \alpha\beta(x-\gamma)^{\beta-1}e^{-\alpha(x-\gamma)^\beta} \qquad \begin{array}{l} x > \gamma \\ \alpha > 0 \\ \beta > 0 \end{array}$$

The implication of this definition of the density is that there is some minimum or "threshold" value γ below which the random variable X cannot fall. In most physical applications this value is 0. For this reason, we shall define the Weibull density with this fact in mind. Be careful when reading scientific literature to note the form of the Weibull density being used.

Definition 4.7.1 (Weibull distribution). A random variable X is said to have a Weibull distribution with parameters α and β if its density is given by

$$f(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^\beta} \qquad \begin{array}{l} x > 0 \\ \alpha > 0 \\ \beta > 0 \end{array}$$

It is easy to verify that the function given in Definition 4.7.1 is a density. (See Exercise 61.) We shall find the mean of this distribution directly rather than by means of the moment generating function.

Theorem 4.7.1. Let X be a Weibull random variable with parameters α and β. The mean and variance of X are given by

$$\mu = \alpha^{-1/\beta}\Gamma(1 + 1/\beta)$$

and

$$\sigma^2 = \alpha^{-2/\beta}\Gamma(1 + 2/\beta) - \mu^2$$

Proof. By Definition 4.2.1

$$E[X] = \int_0^\infty x\alpha\beta x^{\beta-1}e^{-\alpha x^\beta}\,dx$$

$$= \int_0^\infty \alpha\beta x^{\beta}e^{-\alpha x^\beta}\,dx$$

Let $z = \alpha x^\beta$. This implies that

$$x = (z/\alpha)^{1/\beta} \qquad \text{and} \qquad dx = (1/\alpha\beta)(z/\alpha)^{1/\beta-1}\,dz$$

By substitution, it is seen that

$$E[X] = \int_0^\infty \alpha\beta(z/\alpha)e^{-z}(1/\alpha\beta)(z/\alpha)^{1/\beta-1}\,dz$$

$$= \int_0^\infty (z/\alpha)^{1/\beta}e^{-z}\,dz$$

$$= \alpha^{-1/\beta}\int_0^\infty z^{1/\beta}e^{-z}\,dz$$

The integral on the right is, by definition, $\Gamma(1 + 1/\beta)$. (See Definition 4.3.1.) Thus, we have shown that the mean of the Weibull distribution is

$$\mu = E[X] = \alpha^{-1/\beta}\Gamma(1 + 1/\beta)$$

as claimed. The remainder of the proof is outlined as an exercise. (See Exercises 63 and 64.)

The graph of the Weibull density varies depending on the values of α and β. The general shape resembles that of the gamma density with the curve becoming more symmetric as the value of β increases.

Example 4.7.1. Let X be a Weibull random variable with $\beta = 1$. The density for X is

$$f(x) = \alpha e^{-\alpha x} \qquad x > 0 \qquad \alpha > 0$$

Note that this is the density for an *exponential* random variable. That is, the exponential distribution is a special case of the Weibull distribution with $\beta = 1$. By Theorem 4.7.1

$$\mu = \alpha^{-1/\beta}\Gamma(1 + 1/\beta) = (1/\alpha)\Gamma(2) = 1/\alpha \cdot 1! = 1/\alpha$$

$$\sigma^2 = \alpha^{-2/\beta}\Gamma(1 + 2/\beta) - \mu^2$$

$$= 1/\alpha^2\Gamma(3) - (1/\alpha)^2$$

$$= 2/\alpha^2 - 1/\alpha^2 = 1/\alpha^2$$

Note that these results are consistent with those obtained by viewing this random variable as being exponential. (See Exercise 33.)

Reliability

As we have said, the Weibull distribution frequently arises in the study of reliability. Reliability studies are concerned with assessing whether or not a system functions adequately under the conditions for which it was designed. Interest centers on describing the behavior of the random variable X, the time to failure of a system that cannot be repaired once it fails to operate. Three functions come into play when assessing reliability. These are the failure density f, the reliability function R, and ρ, the failure or hazard rate of the distribution. To understand how these functions are defined, consider some system being put into operation at time $t = 0$. We observe the system until it eventually fails. Let X denote the time of the failure. This random variable is continuous and "a priori" can assume any value in the interval $(0, \infty)$. The density f, for X, is called the *failure density* for the component. The *reliability function*, R, is defined to be the probability that the component will not fail before time t. Thus

$$\begin{aligned} R(t) &= 1 - P[\text{component will fail before time } t] \\ &= 1 - \int_0^t f(x)\,dx \\ &= 1 - F(t) \end{aligned}$$

where F is the cumulative distribution function for X. To define ρ, the hazard rate function, consider a time interval $[t, t + \Delta t]$ of length Δt. We define the force of mortality or hazard rate function over this interval by

$$\begin{aligned} \rho(t) &= \lim_{\Delta t \to 0} P(t \le X \le t + \Delta t | t \le X) \frac{1}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\text{probability of failure in } [t, t + \Delta t]}{\text{probability of failure in } [t, \infty)} \cdot \frac{1}{\Delta t} \end{aligned}$$

That is, $\rho(t)$ is the instantaneous rate of failure of the system in the interval $[t, t + \Delta t]$ given that the system is working at time t.

Theorem 4.7.2 relates the three functions f, R, and ρ.

Theorem 4.7.2. Let X be a random variable with failure density f, reliability function R, and hazard rate function ρ. Then

$$\rho(t) = \frac{f(t)}{R(t)}$$

Proof. By definition

$$\begin{aligned} \rho(t) &= \lim_{\Delta t \to 0} \frac{\text{probability of failure in } [t, t + \Delta t]}{\text{probability of failure in } [t, \infty]} \cdot \frac{1}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\int_t^{t+\Delta t} f(x)\,dx}{\int_t^{\infty} f(x)\,dx} \cdot \frac{1}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{1 - F(t)} \cdot \frac{1}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{R(t)} \\ &= \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)} \end{aligned}$$

The job of the scientist is to find the form of these functions for the problem at hand. In practice one often begins by assuming a particular form for the hazard rate function based on empirical evidence. To do so, one must have some practical way to interpret ρ. A rough interpretation is as follows:

1. If ρ is increasing over an interval then, as time goes by, a failure is more likely to occur. This normally happens for systems which begin to fail primarily due to wear.
2. If ρ is decreasing over an interval, then as time goes by a failure is less likely to occur than it was earlier in the time interval. This happens in situations in which defective systems tend to fail early. As time goes by the hazard rate for a well-made system decreases.
3. A steady hazard rate is expected over the useful lifespan of a component. A failure tends to occur during this period due mainly to random factors.

Since one often has an idea of the form only of ρ, the natural question to ask is: "Can we derive the failure density and the reliability function from knowledge of ρ?" Theorem 4.7.3 shows how this can be done.

Theorem 4.7.3. Let X be a random variable with failure density f, reliability function R, and hazard rate ρ. Then

$$R(t) = \exp\left[-\int_0^t \rho(x)\,dx\right]$$

and $f(t) = \rho(t)R(t)$.

Proof. Note that since $R(x) = 1 - F(x)$, $R'(x) = -F'(x)$. Therefore

$$\rho(x) = \frac{f(x)}{R(x)} = \frac{F'(x)}{R(x)} = \frac{-R'(x)}{R(x)}$$

We integrate each side of this equation to obtain

$$\int_0^t \rho(x)\,dx = -\int_0^t \frac{R'(x)}{R(x)}\,dx = -[\ln R(t) - \ln R(0)]$$

Note that $R(0) = 1$ since the component will not fail before time $t = 0$, the moment that it is put into operation. Since $\ln R(0) = \ln 1 = 0$, we see that

$$-\int_0^t \rho(x)\,dx = \ln R(t)$$

or that

$$\exp\left[-\int_0^t \rho(x)\,dx\right] = e^{\ln R(t)} = R(t)$$

as claimed.

Example 4.7.2 illustrates the use of Theorem 4.7.3 and shows how the Weibull distribution arises in reliability studies.

Example 4.7.2. One hazard rate function in widespread use is the function

$$\rho(t) = \alpha\beta t^{\beta-1} \qquad t > 0$$
$$\alpha > 0$$
$$\beta > 0$$

This function has the property that if $\beta = 1$, the hazard rate is constant indicating that the occurrence of a failure is due primarily to random factors; if $\beta > 1$, the hazard rate is increasing, indicating that a failure is due primarily to a system wearing out over time; if $\beta < 1$, the hazard rate is decreasing, indicating that an early failure is likely due to a malfunctioning system. (See Exercise 64.) The reliability function is given by

$$R(t) = \exp\left[-\int_0^t \alpha\beta x^{\beta-1}\,dx\right]$$
$$= \exp\left[-\alpha x^{\beta}\Big|_0^t\right] = e^{-\alpha[t^{\beta}-0^{\beta}]} = e^{-\alpha t^{\beta}}$$

The failure density is given by

$$f(t) = \rho(t)R(t) = \alpha\beta t^{\beta-1}e^{-\alpha t^{\beta}}$$

This is the density for a Weibull random variable with parameters α and β.

Reliability of Series and Parallel Systems

Components in multiple component systems can be installed within the system in various ways. Many systems are arranged in a "series" configuration, some are in "parallel," and others are combinations of the two designs. A system whose components are arranged in such a way that the system fails whenever *any* of its components fail is called a *series* system; if the system fails only if *all* of its components fail, then it is said to be arranged in *parallel*.

Recall that the reliability function for a component is the probability that it will not fail before time t. Consider a system consisting of k components connected in series. Let $R_i(t)$ denote the reliability of component i and assume that the components are independent in the sense that the reliability of one is unaffected by the reliability of the others. The reliability of the entire system is the probability that the system will not fail before time t. The system will not fail if and only if no component fails before time t. Thus the reliability of the system, $R_s(t)$, is given by

$$R_s(t) = \prod_{i=1}^{k} R_i(t)$$

The next two examples illustrate the use of this equation.

Example 4.7.3. Consider a system with five components connected in series. If each component has reliability .95 at time t, then the system reliability at that time is $R_s(t) = (.95)^5 = .774$.

Example 4.7.4. Suppose that we are designing a system of five independent components and we want the system reliability at time t to be at least .95. If the reliability of each component at t is to be the same, what is the minimum reliability required per component? Here we want $x^5 \geq .95$ where x is the reliability of each component. The solution is $x = (.95)^{1/5} = .9898$.

A more practical design for most equipment is the parallel system. Consider k independent components arranged in parallel. When the first fails, the second is used; when the second fails, the third comes on line; this continues until the last component fails, at which time the system fails. The system reliability at time t in this case is the probability that at least one of the k components does not fail before time t. This probability is given by

$$R_s(t) = 1 - P\,(\text{all components fail})$$
$$= 1 - \prod_{i=1}^{k} [1 - R_i(t)]$$

It should be noted that in both series and parallel systems, the reliability of individual components can differ. Example 4.7.5 illustrates a system that makes use of both types of configurations.

Example 4.7.5. Consider a system consisting of eight independent components connected as shown in Fig. 4.11. Note that the system basically consists of five assemblies in series where assembly I consists of component 1; assembly II consists of components 2 and 3 in parallel; assembly III consists of components 4, 5, and 6 in parallel; assemblies IV and V consist of components 7 and 8, respectively. To calculate the system reliability, we first calculate the reliability of the two parallel assemblies. The reliability of assembly II is

$$\left[1 - (1 - .95)^2\right] = .9975$$

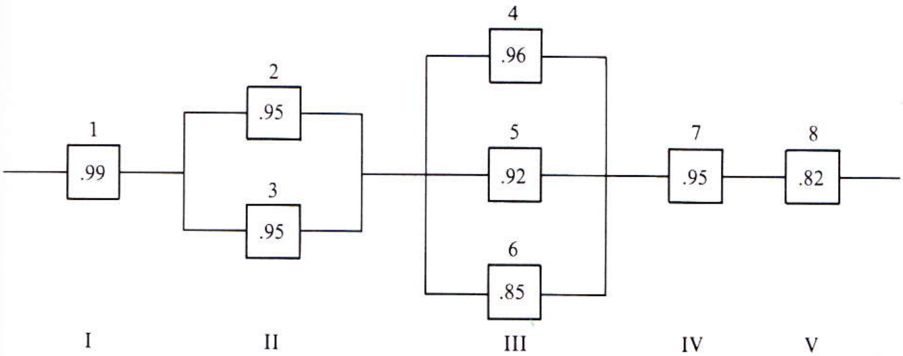


FIGURE 4.11
System with five assemblies with assemblies II and III in parallel.

that of assembly III is

$$[1 - (1 - .96)(1 - .92)(1 - .85)] = .99952$$

The system reliability is the product of the reliabilities of the five assemblies and is given by

$$R_s(t) = (.99)(.9975)(.99952)(.95)(.82) = .7689$$

It is evident that a system with many independent components connected in series may have a very low system reliability even if each component alone is highly reliable. For example, a system of 20 components each with a reliability of .95 connected in series has a system reliability of $(.95)^{20} = .358$. One way to increase system reliability is to replace single components with several similar components arranged in parallel. Of course, the cost of providing this sort of redundancy is usually high.

4.8 TRANSFORMATION OF VARIABLES (OPTIONAL)

Consider a continuous random variable X with density f_X. Suppose that interest centers on some random variable Y where Y is a function of X. Can we determine the density for Y based on knowledge of the distribution of X? The next theorem allows us to answer this question whenever Y is a strictly monotonic function of X.

Theorem 4.8.1. Let X be a continuous random variable with density f_X. Let $Y = g(X)$ where g is strictly monotonic and differentiable. The density for Y is denoted by f_Y and is given by

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$$

Proof. Assume that $Y = g(X)$ is a strictly decreasing function of X. By definition the cumulative distribution function for Y is

$$F_Y(y) = P[Y \le y] = P[g(X) \le y]$$

Since g is strictly decreasing, g^{-1} exists and is also decreasing. Therefore

$$\begin{aligned} P[g(X) \le y] &= P\left[g^{-1}(g(X)) \ge g^{-1}(y)\right] \\ &= P\left[X \ge g^{-1}(y)\right] \\ &= 1 - P\left[X \le g^{-1}(y)\right] \end{aligned}$$

By definition $P[X \le g^{-1}(y)] = F_X(g^{-1}(y))$ and thus substitution yields

$$F_Y(y) = 1 - F_X\left(g^{-1}(y)\right)$$

Since the derivative of the cumulative distribution function yields the density,

$$f_Y(y) = (-1)f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy}$$

Note that since g^{-1} is decreasing, $dg^{-1}(y)/dy < 0$ and

$$-\frac{dg^{-1}(y)}{dy} = \left|\frac{dg^{-1}(y)}{dy}\right|$$

By substitution, $f_Y(y)$ can be written as

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$$

as claimed. The proof in the case in which g is increasing is similar and is left as an exercise. (See Exercise 69.)

An example will illustrate the idea.

Example 4.8.1. Let X be a random variable with density

$$f_X(x) = 2x \qquad 0 < x < 1$$

and let $g(X) = Y = 3X + 6$. Since $g(x) = 3x + 6$ is strictly increasing and differentiable, Theorem 4.8.1 is applicable. To obtain the expression for g^{-1}, we solve the equation $y = 3x + 6$ for x and see that

$$x = g^{-1}(y) = \frac{y-6}{3}$$

and

$$\frac{dg^{-1}(y)}{dy} = \frac{1}{3}$$

An application of Theorem 4.8.1 yields

$$f_Y(y) = f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy}$$

or

$$f_Y(y) = 2\frac{(y-6)}{3}\cdot\frac{1}{3} = \frac{2}{9}(y-6) \qquad 6 < y < 9$$

It should be pointed out that the results given in Theorem 4.8.1 can be applied to piecewise monotonic functions as well as those that are strictly monotonic. In this case, several different equations might be required to define the density for Y. The idea is illustrated in the next example.

Example 4.8.2. Let X be uniformly distributed over $(0,4)$ and let $g(X) = Y = (X-3)^2$. The graph of this function is shown in Fig. 4.12. Note that since g is strictly decreasing on $(0,3)$ and strictly increasing on $(3,4)$, it is piecewise mono-

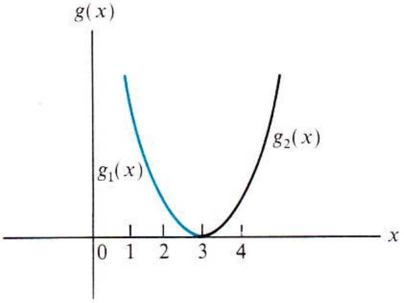


FIGURE 4.12
Graph of $g(x) = (x-3)^2, 0 < x < 4$ partitioned into two monotonic functions $g_1(x) = (x-3)^2, 0 < x \le 3$ and $g_2(x) = (x-3)^2$, $3 < x < 4$.

tonic. It can be defined in terms of the two one-to-one functions g_1 and g_2 given by

$$g(x) = \begin{cases} g_1(x) = (x-3)^2 & 0 < x \le 3 \\ g_2(x) = (x-3)^2 & 3 < x < 4 \end{cases}$$

The functions g_1 and g_2 are each invertible and their inverses are given by

$$g_1^{-1}(y) = 3 - \sqrt{y} \qquad 0 \le y < 9$$

$$g_2^{-1}(y) = 3 + \sqrt{y} \qquad 0 < y < 1$$

Functions h_1 and h_2 used to determine the density for Y are found by applying Theorem 4.8.1 to each of the above. With this done, we then add those functions having common domains to obtain the final expression for f_Y. In this case, the density is formed from the functions

$$h_1(y) = f_X(g_1^{-1}(y))\left|\frac{dg_1^{-1}(y)}{dy}\right| = \frac{1}{4}\left|-\frac{1}{2\sqrt{y}}\right| = \frac{1}{8\sqrt{y}}$$

$$h_2(y) = f_X(g_2^{-1}(y))\left|\frac{dg_2^{-1}(y)}{dy}\right| = \frac{1}{4}\left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{8\sqrt{y}}$$

To obtain the density for Y, we note that the interval $[0,1)$ is common to the domain of both h_1 and h_2. Thus

$$f_Y(y) = h_1(y) + h_2(y) = \frac{2}{8\sqrt{y}} \qquad 0 \le y < 1$$

The interval $[1,9]$ is contained only in the domain of h_2. Hence

$$f_Y(y) = h_2(y) = \frac{1}{8\sqrt{y}} \qquad 1 \le y < 9$$

You can verify for yourself that f_Y is a valid density.

4.9 SIMULATING A CONTINUOUS DISTRIBUTION (OPTIONAL)

In Sec. 3.9 we showed how to simulate a discrete distribution using a random digit table. The table also can be used to simulate a continuous distribution. The

TABLE 4.1
Continuous distributions: a summary

Name	Density		Moment generating function	Mean	Variance
Gamma	$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$	$\alpha > 0$ $\beta > 0$ $x > 0$	$(1 - \beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
Exponential	$\frac{1}{\beta} e^{-x/\beta}$	$x > 0$ $\beta > 0$	$(1 - \beta t)^{-1}$	β	β^2
Chi-squared	$\frac{1}{\Gamma(\gamma/2)2^{\gamma/2}} x^{\gamma/2-1} e^{-x/2}$	$x > 0$ γ a positive integer	$(1 - 2t)^{-\gamma/2}$	γ	2γ
Uniform	$\frac{1}{b - a}$	$a < x < b$	$\frac{e^{tb} - e^{ta}}{t(b - a)}$ $t \neq 0$ 1 $t = 0$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Cauchy	$\frac{1}{\pi} \frac{a}{a^2 + (x - b)^2}$	$-\infty < x < \infty$ $-\infty < b < \infty$ $a > 0$	does not exist	does not exist	does not exist
Normal	$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-1/2\left(\frac{x - \mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$ $\sigma > 0$ $-\infty < \mu < \infty$	$e^{\mu t + \sigma^2 t^2/2}$	μ	σ^2
Weibull	$\alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}$	$x > 0$ $\alpha > 0$ $\beta > 0$		$\alpha^{-1/\beta}\Gamma\left(1 + \frac{1}{\beta}\right)$	$\alpha^{-2/\beta}\Gamma\left(1 + \frac{2}{\beta}\right) - \mu^2$

idea is as follows:

1. We find the cumulative distribution function F for the random variable and its inverse.
2. We select a random two- (or three-) digit number from Table III of App. A and interpret this number as a probability, that is, as a number between 0 and 1.
3. We evaluate F^{-1} at this randomly selected point to obtain a randomly generated value for the random variable X.

This procedure is illustrated in Example 4.9.1.

Example 4.9.1. Consider the random variable X, the time to failure of a computer chip. Assume that X has a Weibull distribution with parameters $\alpha = .02$ and $\beta = 1$. The density for X is

$$f(x) = .02e^{-.02x} \qquad x > 0$$

and its cumulative distribution is

$$y = F(x) = 1 - e^{-.02x}$$

The inverse of F is found by solving this equation for x as follows:

$$\begin{aligned} y &= 1 - e^{-.02x} \\ e^{-.02x} &= 1 - y \\ -.02x &= \ln(1 - y) \\ x &= \frac{-\ln(1 - y)}{.02} \end{aligned}$$

To simulate an observation on X we select a random two-digit number from Table

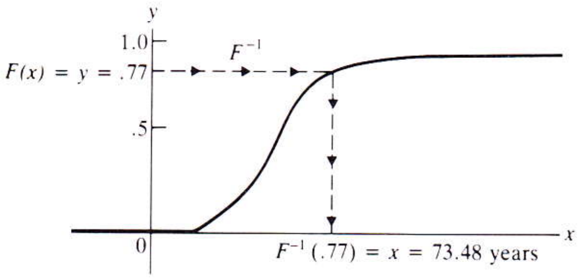


FIGURE 4.13
$F(x) = y = .77$ if and only if $F^{-1}(.77) = x = 73.48$ years.

III of App. A. Suppose the number selected is 77 which is interpreted as the probability $y = .77$. For this value of y, our simulated observation on X is

$$x = \frac{-\ln(1 - .77)}{.02} = 73.48 \text{ years}$$

This procedure can be repeated to generate as many random values for X as desired. Figure 4.13 illustrates this procedure graphically.

CHAPTER SUMMARY

In this chapter we considered the general properties underlying random variables of the continuous type. These are random variables that assume their values in intervals of real numbers rather than an isolated points. The density function was introduced as a means of computing probabilities. These densities are defined in such a way that probabilities correspond to areas. The ideas of expected value and moment generating function were defined by replacing the summation operation, used in the discrete case, with integration. A number of continuous distributions were studied. The gamma distribution was presented. We noted that the exponential distribution and the chi-squared distribution are special cases of the gamma distribution. We studied the normal distribution and showed how to use this distribution to approximate binomial and Poisson probabilities. The Weibull distribution was introduced and its use in reliability studies was examined. Other distributions that were considered briefly are the log-normal, uniform, and Cauchy distributions. We saw how to simulate continuous distributions. These terms were introduced:

Continuous random variable
Continuous distribution function
Half-unit correction
Reliability function
Standard normal
Continuous density
Gamma function
Failure density
Hazard rate function

In the last two chapters, we have presented some commonly encountered discrete and continuous distributions and have looked at some of the relationships that exist among them. The chart given in Fig. 4.14 summarizes the results that have been obtained. It is an adaptation of the more complete chart developed by Lawrence Leemis in "Relationships Among Common Univariate Distributions," *The American Statistician*, May 1986, vol. 40, no. 2. (Used with permission of the author.) In the chart two types of relationships, namely, special cases and approximations, are depicted. Special cases are indicated by a solid arrow and approximations are shown by a dashed arrow. In each case the name of the distribution along with its associated parameters are given.

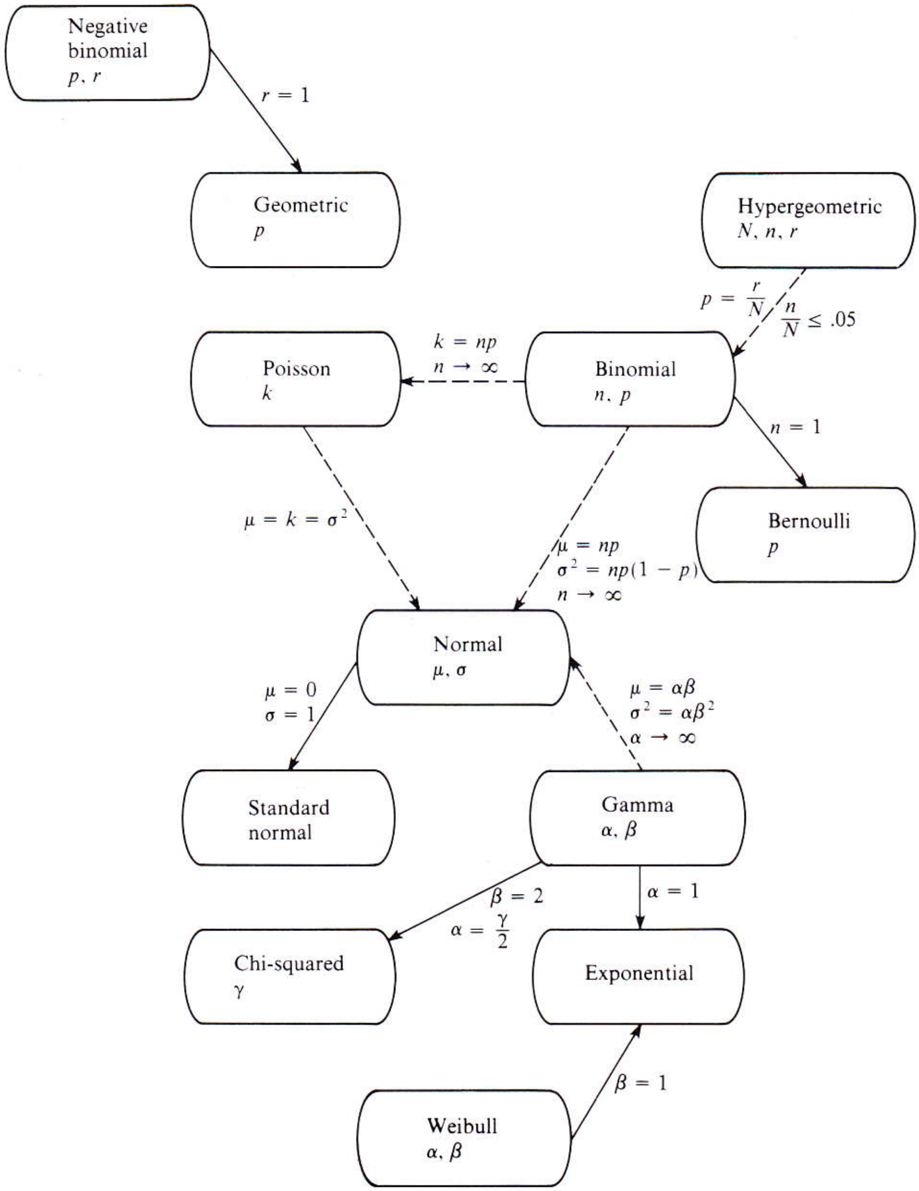


FIGURE 4.14
Some interrelationships among common distributions.

EXERCISES

Section 4.1

1. Consider the function

$$f(x) = kx \qquad 2 \le x \le 4$$

   (*a*) Find the value of k that makes this a density for a continuous random variable.
   (*b*) Find $P[2.5 \le X \le 3]$.
   (*c*) Find $P[X = 2.5]$.
   (*d*) Find $P[2.5 < X \le 3]$.

2. Consider the areas shown in Fig. 4.15. In each case, state what probability is being depicted. What is the relationship between the areas depicted in Figs. 4.15(*a*) and (*b*)? Between those in Fig. 4.15(*d*) and (*e*)?

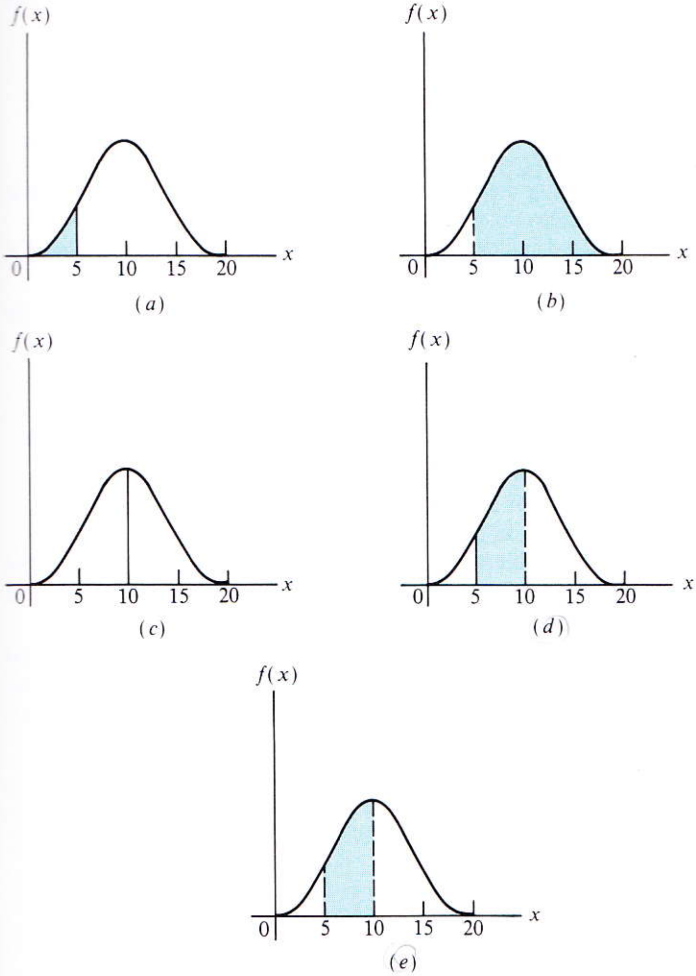


FIGURE 4.15

3. Let X denote the length in minutes of a long-distance telephone conversation. Assume that the density for X is given by

$$f(x) = (1/10)e^{-x/10} \qquad x > 0$$

(*a*) Verify that f is a density for a continuous random variable.
(*b*) Assuming that f adequately describes the behavior of the random variable X, find the probability that a randomly selected call will last at most seven minutes; at least seven minutes; exactly seven minutes.
(*c*) Would it be unusual for a call to last between one and two minutes? Explain, based on the probability of this occurring.
(*d*) Sketch the graph of f and indicate in the sketch the area corresponding to each of the probabilities found in part (*b*).

4. Some plastics in scrapped cars can be stripped out and broken down to recover the chemical components. The greatest success has been in processing the flexible polyurethane cushioning found in these cars. Let X denote the amount of this material, in pounds, found per car. Assume that the density for X is given by

$$f(x) = \frac{1}{\ln 2}\frac{1}{x} \qquad 25 \le x \le 50$$

(Based on a report in *Design Engineering*, February, 1982, p. 7.)
(*a*) Verify that f is a density for a continuous random variable.
(*b*) Use f to find the probability that a randomly selected auto will contain between 30 and 40 pounds of polyurethane cushioning.
(*c*) Sketch the graph of f and indicate in the sketch the area corresponding to the probability found in part (*b*).

5. (*Continuous uniform distribution.*) A random variable X is said to be uniformly distributed over an interval (a, b) if its density is given by

$$f(x) = \frac{1}{b-a} \qquad a < x < b$$

(*a*) Show that this is a density for a continuous random variable.
(*b*) Sketch the graph of the uniform density.
(*c*) Shade the area in the graph of part (*b*) that represents $P[X \le (a+b)/2]$.
(*d*) Find the probability pictured in part (*c*).
(*e*) Let (c, d) and (e, f) be subintervals of (a, b) of equal length. What is the relationship between $P[c \le X \le d]$ and $P[e \le X \le f]$? Generalize the idea suggested by this example thus justifying the name "uniform" distribution.

6. When a pair of coils is placed around a homing pigeon and a magnetic field applied that reverses the earth's field, it is thought that the bird will become disoriented. Under these circumstances it is just as likely to fly in one direction as in any other.

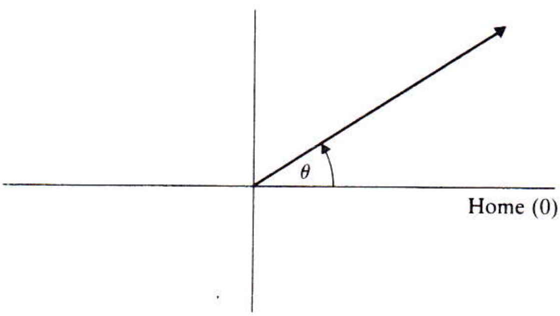


FIGURE 4.16
θ = direction of the initial flight of a homing pigeon measured in radians.

Let θ denote the direction in radians of the bird's initial flight. See Fig. 4.16. θ is uniformly distributed over the interval $[0, 2\pi]$.

(*a*) Find the density for θ.

(*b*) Sketch the graph of the density. The uniform distribution is sometimes called the "rectangular" distribution. Do you see why?

(*c*) Shade the area corresponding to the probability that a bird will orient within $\pi/4$ radians of home, and find this area using plane geometry.

(*d*) Find the probability that a bird will orient within $\pi/4$ radians of home by integrating the density over the appropriate region(s), and compare your answer to that obtained in part (*c*).

(*e*) If 10 birds are released independently and at least seven orient within $\pi/4$ radians of home would you suspect that perhaps the coils are not disorienting the birds to the extent expected? Explain, based on the probability of this occurring.

7. Use Definition 4.1.2 to show that for a continuous random variable X, $P[X = a] = 0$ for every real number a. *Hint:* Write $P[X = a]$ as $P[a \leq X \leq a]$.

8. Express each of the probabilities depicted in Fig. 4.15 in terms of the cumulative distribution function F.

9. Consider the random variable of Exercise 1.

(*a*) Find the cumulative distribution function F.

(*b*) Use F to find $P[2.5 \leq X \leq 3]$ and compare your answer to that obtained previously.

*(*c*) Sketch the graph of F. Is F right continuous? Is F continuous? What is $\lim_{x \to -\infty} F(x)$? What is $\lim_{x \to \infty} F(x)$? Is F nondecreasing?

*(*d*) Find $dF(x)/dx$ for $x \in (2, 4)$. Does your answer look familiar?

10. (*Uniform distribution.*) Find the general expression for the cumulative distribution function for a random variable X that is uniformly distributed over the interval (a, b). See Exercise 5.

11. (*Uniform distribution.*) Consider the random variable of Exercise 6.

(*a*) Use Exercise 10 to find the cumulative distribution function F.

*(*b*) Sketch the graph of F. Is F right continuous? Is F continuous? What is $\lim_{x \to -\infty} F(x)$? What is $\lim_{x \to \infty} F(x)$? Is F nondecreasing?

*(*c*) Find $dF(x)/dx$ for $x \in (0, 2\pi)$. Does your answer look familiar?

12. Find the cumulative distribution function for the random variable of Exercise 3. Use F to find $P[1 \leq X \leq 2]$ and compare your answer to that obtained previously.

13. Find the cumulative distribution function for the random variable of Exercise 4. Use F to find $P[30 \leq X \leq 40]$ and compare your answer to that obtained previously.

***14.** In Exercise 13 of Chap. 3 the mathematical properties of the cumulative distribution function for discrete random variables were pointed out. In the continuous case, similar properties hold. The results of Exercises 9 and 11 are not coincidental! It can be shown that the cumulative distribution function F for any continuous random variable has these characteristics:

(i) F is continuous.

(ii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

(iii) F is nondecreasing.

(iv) $dF(x)/dx = f(x)$ for all values of x for which this derivative exists.

(*a*) Consider the function F defined by

$$F(x) = \begin{cases} 0 & x < -1 \\ x + 1 & -1 \leq x \leq 0 \\ 1 & x > 0 \end{cases}$$

Does F satisfy properties (i) to (iii) given above? If so, what is f? If not, what property fails?

(*b*) Consider the function defined by

$$F(x) = \begin{cases} 0 & x \le 0 \\ x^2 & 0 < x \le 1/2 \\ (1/2)x & 1/2 < x \le 1 \\ 1 & x > 1 \end{cases}$$

Does F satisfy properties (i) to (iii) given above? If so, what is f? If not, what property fails?

Section 4.2

15. Consider the random variable X with density

$$f(x) = (1/6)x \qquad 2 \le x \le 4$$

(*a*) Find $E[X]$.
(*b*) Find $E[X^2]$.
(*c*) Find σ^2 and σ.

16. Let X denote the amount in pounds of polyurethane cushioning found in a car. (See Exercise 4.) The density for X is given by

$$f(x) = \frac{1}{\ln 2}\frac{1}{x} \qquad 25 \le x \le 50$$

Find the mean, variance, and standard deviation for X.

17. Let X denote the length in minutes of a long-distance telephone conversation. The density for X is given by

$$f(x) = (1/10)e^{-x/10} \qquad x > 0$$

(*a*) Find the moment generating function, $m_X(t)$.
(*b*) Use $m_X(t)$ to find the average length of such a call.
(*c*) Find the variance and standard deviation for X.

18. (*Uniform distribution.*) The density for a random variable X distributed uniformly over (a, b) is

$$f(x) = \frac{1}{b-a} \qquad a < x < b$$

Use Definition 4.2.1, to show that

$$E[X] = \frac{a+b}{2} \qquad \text{and} \qquad \operatorname{Var} X = \frac{(b-a)^2}{12}$$

19. (*Uniform distribution.*) Let θ denote the direction in radians of the flight of a bird whose sense of direction has been disoriented as described in Exercise 6. Assume that θ is uniformly distributed over the interval $[0, 2\pi]$. Use the results of Exercise 18 to find the mean, variance, and standard deviation of θ.

20. Let X be continuous with density f. Imagine cutting out of a piece of thin rigid metal the region bounded by the graph of f and the x axis, and attempting to balance this region on a knife-edge held parallel to the vertical axis. The point at which the region would balance, if such a point exists, is the mean of X. Thus μ_X is a "location" parameter in that it indicates the position of the center of the density along the x axis. Figure 4.17 gives the graphs of the densities of four continuous

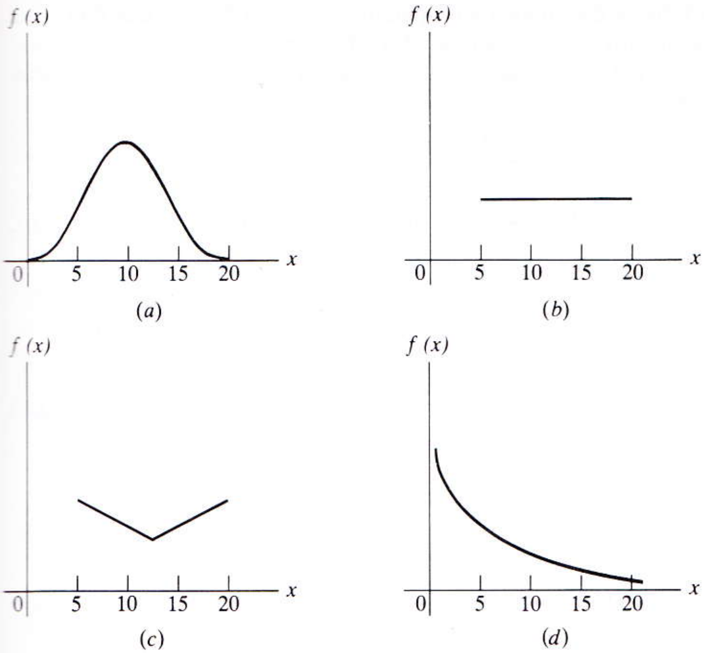


FIGURE 4.17

random variables whose means do exist. In each case, approximate the value of μ_X from the graph.

21. In the continuous case variance is a "shape" parameter in the sense that a random variable with small variance will have a compact density; one with a large variance will have a density that is rather spread out or flat. Consider the two densities given in Fig. 4.18. What is μ_X? What is μ_Y? Which random variable has the larger variance?

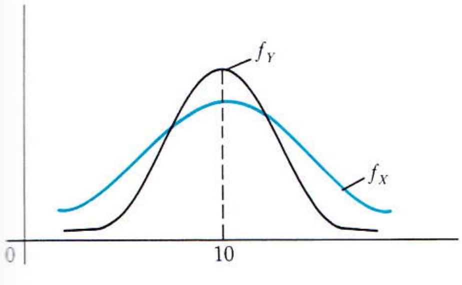


FIGURE 4.18

***22.** (*Cauchy distribution.*) A random variable X with density

$$f(x) = \frac{1}{\pi} \frac{a}{a^2 + (x - b)^2} \qquad \begin{array}{c} -\infty < x < \infty \\ -\infty < b < \infty \\ a > 0 \end{array}$$

is said to have a Cauchy distribution with parameters a and b. This distribution is interesting in that it provides an example of a continuous random variable whose mean does not exist. Let $a = 1$ and $b = 0$ to obtain a special case of the Cauchy distribution with density

$$f(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \qquad -\infty < x < \infty$$

Show that $\int_{-\infty}^{\infty}|x|f(x)\,dx$ does not exist thus showing that $E[X]$ does not exist, *Hint:* Write

$$\int_{-\infty}^{\infty}|x|\frac{1}{\pi}\frac{1}{1+x^2}\,dx = \frac{1}{\pi}\int_{-\infty}^{0}\frac{-x}{1+x^2}\,dx + \frac{1}{\pi}\int_{0}^{\infty}\frac{x}{1+x^2}\,dx$$

and recall that $\int(du/u) = \ln|u|$.

***23.** (*Uniform distribution.*) Let X be uniformly distributed over (a, b). (See Exercise 18.)

(*a*) Show that the moment generating function for X is given by

$$m_X(t) = \begin{cases} \dfrac{e^{tb} - e^{ta}}{t(b-a)} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

Hint: When $t = 0$, $m_X(t) = E[e^{0\cdot X}]$.

(*b*) Use $m_X(t)$ to find $E[X]$. *Hint:* Find

$$\frac{d}{dt}\left(\frac{e^{tb} - e^{ta}}{t(b-a)}\right)$$

and take the limit of this derivative as $t \to 0$ using L'Hospital's rule.

***24.** Let the density for X be given by

$$f(x) = ce^{-|x|} \qquad -\infty < x < \infty$$

(*a*) Find the value of c that makes this a density.

(*b*) Show that

$$\frac{1}{2}\int_{-\infty}^{\infty}|x|e^{-|x|}\,dx$$

exists.

(*c*) Find $E[X]$.

(*d*) Show that

$$m_X(t) = \frac{-1}{t^2 - 1} \qquad -1 < t < 1$$

(*e*) Use $m_X(t)$ to find $E[X]$ and $E[X^2]$

(*f*) Find Var X.

Section 4.3

25. Evaluate each of these integrals:

(*a*) $\int_0^\infty z^2e^{-z}\,dz$

(*b*) $\int_0^\infty z^7e^{-z}\,dz$

(*c*) $\int_0^\infty x^3e^{-x/2}\,dx$

(*d*) $\int_0^\infty (1/16)xe^{-x/4}\,dx$

26. Prove Theorem 4.3.1. *Hint:* To prove part 1, evaluate $\Gamma(1)$ directly from the definition of the gamma function. To prove part 2, use integration by parts with

$$u = z^{\alpha-1} \qquad dv = \int e^{-z}\, dz$$

$$du = (\alpha - 1) z^{\alpha-2}\, dz \qquad v = -e^{-z}$$

Use L'Hospital's rule repeatedly to show that

$$-z^{\alpha-1}e^{-z}\Big|_0^\infty = 0$$

27. (*a*) Use Theorem 4.3.1 to evaluate $\Gamma(2)$, $\Gamma(3)$, $\Gamma(4)$, $\Gamma(5)$, and $\Gamma(6)$.
(*b*) Can you generalize the pattern suggested in part *a*?
(*c*) Does the result of part *b* hold even if $n = 1$?
(*d*) Evaluate $\Gamma(15)$ using the result of part *b*.

28. Show that for $\alpha > 0$ and $\beta > 0$

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}\, dx = 1$$

thereby showing that the function given in Definition 4.3.2 is a density for a continuous random variable. *Hint:* Change the variable by letting $z = x/\beta$.

29. Let X be a gamma random variable with $\alpha = 3$ and $\beta = 4$.
(*a*) What is the expression for the density for X?
(*b*) What is the moment generating function for X?
(*c*) Find μ, σ^2, and σ.

30. Let X be a gamma random variable with parameters α and β. Use the moment generating function to find $E[X]$ and $E[X^2]$. Use these expectations to show that $\operatorname{Var} X = \alpha\beta^2$.

31. Let X be a gamma random variable with parameters α and β.
(*a*) Use Definition 4.2.1, the definition of expected value, to find $E[X]$ and $E[X^2]$ directly. *Hint:* $z^\alpha = z^{(\alpha+1)-1}$ and $z^{\alpha+1} = z^{(\alpha+2)-1}$.
(*b*) Use the results of (*a*) to verify that $\operatorname{Var} X = \alpha\beta^2$.

32. Show that the graph of the density for a gamma random variable with parameters α and β assumes its maximum value at $x = \beta(\alpha - 1)$ for $\alpha > 1$. Sketch a rough graph of the density for a gamma random variable with $\alpha = 3$ and $\beta = 4$. *Hint:* Find the first derivative of the density, set this derivative equal to 0 and solve for x.

33. Let X be an exponential random variable with parameter β. Find general expressions for the moment generating function, mean, and variance for X.

34. A particular nuclear plant releases a detectable amount of radioactive gases twice a month on the average. Find the probability that at least three months will elapse before the release of the first detectable emission. What is the average time that one must wait to observe the first emission?

35. California is hit by approximately 500 earthquakes that are large enough to be felt every year. However, those of destructive magnitude occur on the average once a year.
(*a*) Find the probability that at least three months elapse before the first earthquake of destructive magnitude occurs. (See Exercise 62, Chap. 3.)
(*b*) Suppose that no destructive quake has occurred for four months. Find the probability that an additional three months will elapse before a destructive

quake occurs. *Hint:* You are asked to find $P[W > 7 | W > 4]$. Use the definition of conditional probability, Definition 2.2.1, to conclude that

$$P[W > 7 | W > 4] = P[W > 7]/P[W > 4].$$

(*c*) Are the answers to parts (*a*) and (*b*) the same?

36. Rock noise in an underground mine occurs at an average rate of three per hour. (See Exercise 63, Chap. 3.)
(*a*) Find the probability that no rock noise will be recorded for at least 30 minutes.
(*b*) Suppose that no rock noise has been heard for 15 minutes. Find the probability that another 30 minutes will elapse before the first rock noise is detected.
(*c*) Are the answers to parts (*a*) and (*b*) the same?

37. The results of Exercise 35 and 36 are not coincidental. They illustrate the "forgetfulness" property of the exponential distribution. This property says that the probability that we must wait a total of $w_1 + w_2$ units before the occurrence of an event given that we have already waited w_1 units is the same as the probability that we must wait w_2 units at the outset. That is,

$$P[W > w_1 + w_2 | W > w_1] = P[W > w_2]$$

Verify this statement for any exponential random variable W with parameter β.

38. Consider a chi-squared random variable with 15 degrees of freedom.
(*a*) What is the mean of X_{15}^2? What is its variance?
(*b*) What is the expression for the density for X_{15}^2?
(*c*) What is the expression for the moment generating function for X_{15}^2?
(*d*) Use Table IV of App. A to find each of the following:

$$P\left[X_{15}^2 \le 5.23\right] \qquad P\left[6.26 \le X_{15}^2 \le 27.5\right] \qquad \chi_{.05}^2$$

$$P\left[X_{15}^2 \ge 22.3\right] \qquad \chi_{.01}^2 \qquad \chi_{.95}^2$$

Section 4.4

39. Use Table V of App. A to find each of the following:
(*a*) $P[Z \le 1.57]$. (*b*) $P[Z < 1.57]$.
(*c*) $P[Z = 1.57]$. (*d*) $P[Z > 1.57]$.
(*e*) $P[-1.25 \le Z \le 1.75]$. (*f*) $z_{.10}$.
(*g*) $z_{.90}$.
(*h*) The point z such that $P[-z \le Z \le z] = .95$.
(*i*) The point z such that $P[-z \le Z \le z] = .90$.

40. The bulk density of soil is defined as the mass of dry solids per unit bulk volume. A high bulk density implies a compact soil with few pores. Bulk density is an important factor in influencing root development, seedling emergence, and aeration. Let X denote the bulk density of Pima clay loam. Studies show that X is normally distributed with $\mu = 1.5$ and $\sigma = .2$ g/cm^3. (*McGraw-Hill Yearbook of Science and Technology*, 1981, p. 361.)
(*a*) What is the density for X? Sketch a graph of the density function. Indicate on this graph the probability that X lies between 1.1 and 1.9. Find this probability.
(*b*) Find the probability that a randomly selected sample of Pima clay loam will have bulk density less than .9 g/cm^3.
(*c*) Would you be surprised if a randomly selected sample of this type of soil has a bulk density in excess of 2.0 g/cm^3? Explain, based on the probability of this occurring.

(*d*) What point has the property that only 10% of the soil samples have bulk density this high or higher?

(*e*) What is the moment generating function for X?

41. Most galaxies take the form of a flattened disc with the major part of the light coming from this very thin fundamental plane. The degree of flattening differs from galaxy to galaxy. In the Milky Way Galaxy, most gases are concentrated near the center of the fundamental plane. Let X denote the perpendicular distance from this center to a gaseous mass. X is normally distributed with mean 0 and standard deviation 100 parsecs. (A parsec is equal to approximately 19.2 trillion miles.) (*McGraw-Hill Encyclopedia of Science and Technology*, vol. 6, 1971, p. 10.)

(*a*) Sketch a graph of the density for X. Indicate on this graph the probability that a gaseous mass is located within 200 parsecs of the center of the fundamental plane. Find this probability.

(*b*) Approximately what percentage of the gaseous masses are located more than 250 parsecs from the center of the plane?

(*c*) What distance has the property that 20% of the gaseous masses are at least this far from the fundamental plane?

(*d*) What is the moment generating function for X?

42. Among diabetics, the fasting blood glucose level X may be assumed to be approximately normally distributed with mean 106 mg/100 ml and standard deviation 8 mg/100 ml.

(*a*) Sketch a graph of the density for X. Indicate on this graph the probability that a randomly selected diabetic will have a blood glucose level between 90 and 122 mg/100 ml. Find this probability.

(*b*) Find $P[X \leq 120 \text{ mg/100 ml}]$.

(*c*) Find the point that has the property that 25% of all diabetics have a fasting glucose level of this value or lower.

(*d*) If a randomly selected diabetic is found to have fasting blood glucose level in excess of 130, do you think there is cause for concern? Explain, based on the probability of this occurring naturally.

43. (*a*) Find the density for the standard normal random variable Z.

(*b*) Find $f'(z)$. Show that the only critical point for f occurs at $z = 0$. Use the first derivative test to show that f assumes its maximum value at $z = 0$.

(*c*) Find $f''(z)$. Show that the possible inflection points occur at $z = \pm 1$. Use the second derivative to show that f changes concavity at $z = \pm 1$ implying that the inflection points do occur when $z = \pm 1$.

(*d*) Let X be normal with parameters μ and σ. Let $(X - \mu)/\sigma = Z$. Use the results of parts (*b*) and (*c*) to verify that, in general, a normal curve assumes its maximum value at $x = \mu$ and has points of inflection at $x = \mu \pm \sigma$.

44. Let X be normal with parameters μ and σ. Use the moment generating function to find $E[X^2]$. Find Var X thus completing the proof of Theorem 4.4.2.

***45.** (*Log-normal distribution.*) The log-normal distribution is the distribution of a random variable whose natural logarithm follows a normal distribution. Thus if X is a normal random variable then $Y = e^X$ follows a log-normal distribution. Complete the argument below thus deriving the density for a log-normal random variable.

Let X be normal with mean μ and variance σ^2. Let G denote the cumulative distribution function for $Y = e^X$ and let F denote the cumulative distribution function for X.

(a) Show that $G(y) = F(\ln y)$

(b) Show that $G'(y) = F'(\ln y)/y$

(c) Use Exercise 14 to show that the density for Y is given by

$$g(y) = \frac{1}{\sqrt{2\pi}\,\sigma y} \exp\left[-(1/2)\frac{(\ln y - \mu)^2}{\sigma}\right] \qquad \begin{array}{l} -\infty < \mu < \infty \\ \sigma > 0 \\ \\ y > 0 \end{array}$$

Note that μ and σ are the mean and standard deviation of the underlying normal distribution; they are not the mean and standard deviation of Y itself.

***46.** Let Y denote the diameter in millimeters of Styrofoam pellets used in packing. Assume that Y has a log-normal distribution with parameters $\mu = .8$ and $\sigma = .1$.

(a) Find the probability that a randomly selected pellet has a diameter that exceeds 2.7 mm.

(b) Between what two values will Y fall with probability approximately .95?

Section 4.5

47. Verify the normal probability rule.

48. The number of Btu's of petroleum and petroleum products used per person in the United States in 1975 was normally distributed with mean 153 million Btu and standard deviation 25 million Btu. Approximately what percentage of the population used between 128 and 178 million Btu during that year? Approximately what percentage of the population used in excess of 228 million Btu?

49. Reconsider Exercises 40(a), 41(a), and 42(a) in light of the normal probability rule.

50. For a normal random variable, $P[|X - \mu| < 3\sigma] \doteq .99$. What value is assigned to this probability via Chebyshev's inequality? Are the results consistent? Which rule gives a stronger statement in the case of a normal variable?

51. Animals have an excellent spatial memory. In an experiment to confirm this statement an eight-armed maze such as that shown in Fig. 4.19 is used. At the

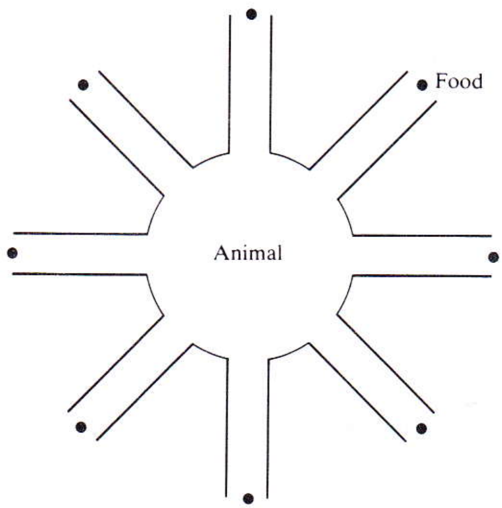


FIGURE 4.19
An eight-armed maze.

beginning of a test, one pellet of food is placed at the end of each arm. A hungry animal is placed at the center of the maze and is allowed to choose freely from among the arms. The optimal strategy is to run to the end of each arm exactly once. This requires that the animal remember where it has been. Let X denote the number of correct arms (arms still containing food) selected among its first eight choices. Studies indicate that $\mu = 7.9$. (*McGraw-Hill Encyclopedia of Science and Technology*, 1980.)

(a) Is X normally distributed?

(b) State and interpret Chebyshev's inequality in the context of this problem for $k = .5$, 1, 2, and 3. At what point does the inequality begin to give us some practical information?

Section 4.6

52. Let X be binomial with $n = 20$ and $p = .3$. Use the normal approximation to approximate each of the following. Compare your results with the values obtained from Table I of App. A.

(*a*) $P[X \leq 3]$.

(*b*) $P[3 \leq X \leq 6]$.

(*c*) $P[X \geq 4]$.

(*d*) $P[X = 4]$.

53. Although errors are likely when taking measurements from photographic images, these errors are often very small. For sharp images with negligible distortion, errors in measuring distances are often no larger than .0004 inches. Assume that the probability of a serious measurement error is .05. A series of 150 independent measurements are made. Let X denote the number of serious errors made.

(*a*) In find the probability of making at least one serious error, is the normal approximation appropriate? If so, approximate the probability using this method.

(*b*) Approximate the probability that at most three serious errors will be made.

54. A chemical reaction is run in which the usual yield is 70%. A new process has been devised that should improve the yield. Proponents of the new process claim that it produces better yields than the old process more than 90% of the time. The new process is tested 60 times. Let X denote the number of trials in which the yield exceeds 70%.

(*a*) If the probability of an increased yield is .9, is the normal approximation appropriate?

(*b*) If $p = .9$, what is $E[X]$?

(*c*) If $p > .9$ as claimed then, on the average, more than 54 of every 60 trials will result in an increased yield. Let us agree to accept the claim if X is at least 59. What is the probability that we will accept the claim if p is really only .9?

(*d*) What is the probability that we will not accept the claim ($X \leq 58$) if it is true, and p is really .95?

55. Opponents of a nuclear power project claim that the majority of those living near a proposed site are opposed to the project. To justify this statement, a random sample of 75 residents is selected and their opinions sought. Let X denote the number opposed to the project.

(*a*) If the probability that an individual is opposed to the project is .5, is the normal approximation appropriate?

(*b*) If $p = .5$, what is $E[X]$?

(*c*) If $p > .5$ as claimed then, on the average, more than 37.5 of every 75 individuals are opposed to the project. Let us agree to accept the claim if X is at least 46. What is the probability that we will accept the claim if p is really only .5?

(*d*) What is the probability that we will not accept the claim ($X \leq 45$) even though it is true and p is really .7?

56. (*Normal approximation to the Poisson distribution.*) Let X be Poisson with parameter λs. Then, for large values of λs, X is approximately normal with mean λs and variance λs. (The proof of this theorem is also based on the Central Limit Theorem and will be considered in Chap. 7.) Let X be a Poisson random variable with parameter $\lambda s = 15$. Find $P[X \leq 12]$ from Table II of App. A. Approximate this probability using a normal curve. Be sure to employ the half-unit correction factor.

57. The average number of jets either arriving at or departing from O'Hare Airport is one every 40 seconds. What is the approximate probability that at least 75 such flights will occur during a randomly selected hour? What is the probability that fewer than 100 such flights will take place in an hour?

Section 4.7

58. The length of time in hours that a rechargeable calculator battery will hold its charge is a random variable. Assume that this variable has a Weibull distribution with $\alpha = .01$ and $\beta = 2$.

(*a*) What is the density for X?

(*b*) What is the mean and variance for X? *Hint:* It can be shown that $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for any $\alpha > 1$. Furthermore $\Gamma(1/2) = \sqrt{\pi}$.

(*c*) What is the reliability function for this random variable?

(*d*) What is the reliability of such a battery at $t = 3$ hours? At $t = 12$ hours? At $t = 20$ hours?

(*e*) What is the hazard rate function for these batteries?

(*f*) What is the failure rate at $t = 3$ hours? At $t = 12$ hours? At $t = 20$ hours?

(*g*) Is the hazard rate function an increasing or a decreasing function? Does this seem to be reasonable from a practical point of view? Explain.

59. Computer chips do not "wear out" in the ordinary sense. Assuming that defective chips have been removed from the market by factory inspection, it is reasonable to assume that these chips exhibit a constant hazard rate. Let the hazard rate be given by $\rho(t) = .02$. (Time is in years.)

(*a*) In a practical sense, what are the main causes of failure of these chips?

(*b*) What is the reliability function for chips of this type?

(*c*) What is the reliability of a chip 20 years after it has been put into use?

(*d*) What is the failure density for these chips?

(*e*) What type of random variable is X, the time to failure of a chip?

(*f*) What is the mean and variance for X?

(*g*) What is the probability that a chip will be operable for at least 30 years?

60. The random variable X, the time to failure (in thousands of miles driven) of the signal lights on an automobile has a Weibull distribution with $\alpha = .04$ and $\beta = 2$.

(*a*) Find the density, mean, and variance for X.

(*b*) Find the reliability function for X.

(*c*) What is the reliability of these lights at 5000 miles? At 10,000 miles?

(*d*) What is the hazard rate function?

(*e*) What is the hazard rate at 5000 miles? At 10,000 miles?

(*f*) What is the probability that the lights will fail during the first 3000 miles driven?

61. Show that for $\alpha > 0$ and $\beta > 0$

$$\int_0^\infty \alpha\beta x^{\beta-1}e^{-\alpha x^\beta}\,dx = 1$$

thereby showing that the nonnegative function given in Definition 4.7.1 is a density for a continuous random variable. *Hint:* Let $z = \alpha x^\beta$.

62. Let X be a Weibull random variable with parameters α and β. Show that $E[X^2] = \alpha^{-2/\beta}\Gamma(1 + 2/\beta)$. *Hint:* In evaluating

$$\int_0^\infty x^2\alpha\beta x^{\beta-1}e^{-\alpha x^\beta}\,dx$$

let $z = \alpha x^\beta$. Evaluate the integral in a manner similar to that used in the proof of Theorem 4.7.1.

63. Use the result of Exercise 62 to find Var X for a Weibull random variable with parameters α and β thus completing the proof of Theorem 4.7.1.

64. Consider the hazard rate function

$$\rho(t) = \alpha\beta t^{\beta-1} \qquad \begin{aligned} t &> 0\\ \alpha &> 0\\ \beta &> 0\end{aligned}$$

(*a*) Show that $\rho(t)$ is constant if $\beta = 1$.

(*b*) Find $\rho'(t)$. Argue that $\rho'(t) > 0$ if $\beta > 1$ thus producing an increasing hazard rate. Argue that $\rho'(t) < 0$ if $\beta < 1$ thus producing a decreasing hazard rate.

65. A system has eight components connected as shown in Fig. 4.20.

(*a*) Find the reliability of each of the parallel assemblies.

(*b*) Find the system reliability.

(*c*) Suppose that assembly II is replaced by two identical components in parallel each with reliability .98. What is the reliability of the new assembly?

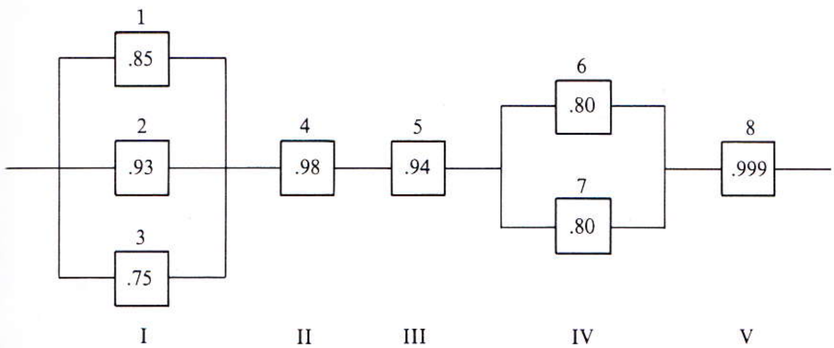


FIGURE 4.20

(d) What is the new system reliability after making the change suggested in part (c)?

(e) Make changes analogous to that of part (c) in each of the remaining single component assemblies. Compute the new system reliability.

66. A system consists of two independent components connected in series. The lifespan of the first component follows a Weibull distribution with $\alpha = .006$ and $\beta = .5$; the second has a lifespan that follows the Weibull distribution with $\beta = 1$ and $\alpha = .00004$.

(a) Find the reliability of the system at 2500 hours.

(b) Find the probability that the system will fail before 2000 hours.

(c) If the two components are connected in parallel, what is the system reliability at 2500 hours?

67. Suppose that a missile can have several independent and identical computers each with reliability .9 connected in parallel so that the system will continue to function as long as at least one computer is operating. If it is desired to have a system reliability of at least .999, how many computers should be connected in parallel?

68. Three independent and identical components, each with a reliability of .9, are to be used in an assembly.

(a) The assembly will function if at least one of the components is operable. Find the system reliability.

(b) The assembly will function if at least two of the components is operable. Find the reliability of the system.

(c) The assembly will function only if all three of the components are operable. Find the reliability of the system.

Section 4.8

69. Prove Theorem 4.8.1 in the case in which g is strictly increasing.

70. Let X be a random variable with density

$$f_X(x) = (1/4)x \qquad 0 \le x \le \sqrt{8}$$

and let $Y = X + 3$.

(a) Find $E[X]$ and then use the rules for expectation to find $E[Y]$.

(b) Find the density for Y.

(c) Use the density for Y to find $E[Y]$ and compare your answer to that found in part (a).

71. Let X be a random variable with density

$$f_X(x) = (1/4)xe^{-x/2} \qquad x \ge 0$$

and let $Y = (-1/2)X + 2$. Find the density for Y.

72. Let X be a random variable with density

$$f_X(x) = e^{-x} \qquad x > 0$$

and let $Y = e^X$. Find the density for Y.

73. Let C denote the temperature in degrees Celsius to which a computer will be subjected in the field. Assume that C is uniformly distributed over the interval $(15, 21)$. Let F denote the field temperature in degrees Fahrenheit so that $F = (9/5)C + 32$. Find the density for F.

74. Let X denote the velocity of a random gas molecule. According to the Maxwell-Boltzmann law, the density for X is given by

$$f_X(x) = cx^2e^{-\beta x^2} \qquad x > 0$$

Here c is a constant that depends on the gas involved and β is a constant whose value depends on the mass of the molecule and its absolute temperature. The kinetic energy of the molecule, Y, is given by $Y = (1/2)mX^2$ where $m > 0$. Find the density for Y.

75. Let X be a continuous random variable with density f_X and let $Y = X^2$.
(a) Show that for $y \geq 0$,
$$F_Y(y) = P\left[-\sqrt{y} \leq X \leq \sqrt{y}\right]$$
(b) Show that for $y \geq 0$,
$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$
(c) Use the technique given in the proof of Theorem 4.8.1 to show that
$$f_Y(y) = 1/(2\sqrt{y})\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right]$$
(d) Use the technique illustrated in Example 4.8.2 to show that
$$f_Y(y) = 1/(2\sqrt{y})\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right]$$

76. Let Z be a standard normal random variable and let $Y = Z^2$.
(a) Show that $\Gamma(1/2) = \int_0^\infty x^{-1/2}e^{-x}\,dx$.
(b) Show that $\Gamma(1/2) = \sqrt{\pi}$. *Hint:* Use the results of part (a) with $x = t^2/2$ and make use of the fact that the standard normal density integrates to 1 when integrated over the set of real numbers.
(c) Use the results of Exercise 75 to find f_Y.
(d) Argue that Y follows a chi-squared distribution with 1 degree of freedom.

77. Let X be normally distributed with mean μ and variance σ^2. Let $Y = e^X$. Show that Y follows the log-normal distribution. (See Exercise 45.)

78. Let Z be a standard normal random variable and let $Y = 2Z^2 - 1$. Find the density for Y.

Section 4.9

79. Use Table III of App. A to generate nine more observations on the random variable X, the time to failure of a computer chip. See Example 4.9.1. Based on these data, approximate the average time to failure by finding the arithmetic average of the values of X simulated in the experiment. Does this value agree well with the theoretical mean value of 50 years?

80. Simulate 20 observations on the random variable X, the time to failure of the signal lights on an automobile. (See Exercise 60.) Approximate the average time to failure for these lights based on the simulated data. Does this value agree well with the theoretical mean value for X?

81. A satellite has malfunctioned and is expected to reenter the earth's atmosphere sometime during a four-hour period. Let X denote the time of reentry. Assume that X is uniformly distributed over the interval $[0, 4]$. Simulate 20 observations on X. (See Exercise 18.)

REVIEW EXERCISES

82. Let X be a continuous random variable with density
$$f(x) = cx^2 \qquad -3 \leq x \leq 3$$
(a) Assuming that $f(x) = 0$ elsewhere, find the value of c that makes this a density.

(*b*) Find $E[X]$ and $E[X^2]$ from the definitions of these terms.
(*c*) Find Var X and σ.
(*d*) Find $P[X \le 2]$; $P[-1 \le X \le 2]$; $P[X > 1]$ by direct integration.
(*e*) Find the closed form expression for the cumulative distribution function F.
(*f*) Use F to find each of the probabilities of part (*d*) and compare your answers to those obtained earlier.

83. Find $\int_0^\infty z^{10} e^{-z}\, dz$.

84. A computer firm introduces a new home computer. Past experience shows that the random variable X, the time of peak demand measured in months after its introduction, follows a gamma distribution with variance 36.
(*a*) If the expected value of X is 18 months, find α and β.
(*b*) Find $P[X \le 7.01]$; $P[X \ge 26]$; $P[13.7 \le X \le 31.5]$.

85. Let X denote the lag time in a printing queue at a particular computer center. That is, X denotes the difference between the time that a program is placed in the queue and the time at which printing begins. Assume that X is normally distributed with mean 15 minutes and variance 25.
(*a*) Find the expression for the density for X.
(*b*) Find the probability that a program will reach the printer within three minutes of arriving in the queue.
(*c*) Would it be unusual for a program to stay in the queue between 10 and 20 minutes? Explain, based on the approximate probability of this occurring. You don't have to use the Z table to answer this question!
(*d*) Would you be surprised if it took longer than 30 minutes for the program to reach the printer? Explain, based on the probability of this occurring.

86. A computer center maintains a telephone consulting service to troubleshoot for its users. The service is available from 9 a.m. to 5 p.m. each working day. Past experience shows that the random variable X, the number of calls received per day, follows a Poisson distribution with $\lambda = 50$. For a given day, find the probability that the first call of the day will be received by 9:15 a.m.; after 3 p.m.; between 9:30 a.m. and 10 a.m.

87. Let $H(X) = X^2 + 3X + 2$. Find $E[H(X)]$ if
(*a*) X is normally distributed with mean 3 and variance 4.
(*b*) X has a gamma distribution with $\alpha = 2$ and $\beta = 4$.
(*c*) X has a chi-squared distribution with 10 degrees of freedom.
(*d*) X has an exponential distribution with $\beta = 5$.
(*e*) X has a Weibull distribution with $\alpha = 2$ and $\beta = 1$.

88. Let the density for the continuous random variable X be given by

$$f(x) = 1/2e^{-|x|} \qquad -\infty < x < \infty$$

(*a*) Show that $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
(*b*) Show that

$$m_X(t) = (1/2)[1/(t+1) - 1/(t-1)] \qquad -1 < t < 1$$

(*c*) Use $m_X(t)$ to show that $E[X] = 0$.

89. Let X denote the time to failure in years of a telephone modem used to access a mainframe computer from a remote terminal. Assume that the hazard rate function for X is given by

$$\rho(t) = \alpha\beta t^{\beta-1}$$

where $\alpha = 2$ and $\beta = 1/5$.

(a) Find the failure density for X.
(b) Find the expected value of X.
(c) Find the reliability function for X.
(d) Find the probability that the modem will last for at least two years.
(e) What is the hazard rate at $t = 1$ year?
(f) Describe roughly the theoretical pattern in the causes of failure in these modems.

90. Past evidence shows that when a customer complains of an out-of-order phone there is an 8% chance that the problem is with the inside wiring. During a one-month period, 100 complaints are lodged. Assume that there have been no wide-scale problems that could be expected to affect many phones at once, and that, for this reason, these failures are considered to be independent. Find the expected number of failures due to a problem with the inside wiring. Find the probability that at least 10 failures are due to a problem with the inside wiring. Would it be unusual if at most five were due to problems with the inside wiring? Explain, based on the probability of this occurring.

91. The cumulative distribution function for a continuous random variable X is defined by

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^3 + x^2}{2} & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Find the density for X.

92. The density for a continuous random variable is given by

$$f(x) = xe^{-x} \qquad 0 < x < \infty$$

(a) Show that $\int_0^\infty xe^{-x}\,dx = 1$. *Hint:* Use the gamma function.
(b) Find $E[X]$, $E[X^2]$ and Var X.
(c) Show that $m_X(t) = 1/(1-t)^2$ where $t < 1$.
(d) Use $m_X(t)$ to find $E[X]$.

93. An electronic counter records the number of vehicles exiting the interstate at a particular point. Assume that the average number of vehicles leaving in a five-minute period is 10. Approximate the probability that between 100 and 120 vehicles inclusive will exist at this point in a one-hour period.

CHAPTER 5

JOINT DISTRIBUTIONS

Thus far, interest has centered on a single random variable of either the discrete or the continuous type. Such random variables are called *univariate*. Problems do arise in which two random variables are to be studied simultaneously. For example, we might wish to study the yield of a chemical reaction in conjunction with the temperature at which the reaction is run. Typical questions to ask are: "Is the yield independent of the temperature?" or: "What is the average yield if the temperature is 40°C?" To answer questions of this type, we need to study what are called *two-dimensional or bivariate random variables* of both the discrete and continuous type. In this chapter we present a brief introduction to the basic theoretical concepts underlying these variables. These concepts form the basis for the study of regression analysis and correlation, topics of extreme importance in applied statistics. (See Chaps. 11 and 12.)

5.1 JOINT DENSITIES AND INDEPENDENCE

We begin by considering two-dimensional random variables and their density functions. The definitions presented here are natural extensions of those presented for a single random variable in Chaps. 3 and 4. (See Definition 3.2.1 and 4.1.2.)

Definition 5.1.1 (Discrete joint density). Let X and Y be discrete random variables. The ordered pair (X, Y) is called a two-dimensional discrete random variable. A function f_{XY} such that

$$f_{XY}(x, y) = P[X = x \text{ and } Y = y]$$

is called the joint density for (X, Y).

Again let us point out that in the discrete case some statisticians prefer to use the term probability function or probability mass function rather than the term "density." We shall use the term density and the notation f_{XY} in both the discrete and the continuous cases for consistency of notation and terminology.

Note that the purpose of the density here is the same as in the past—to allow us to compute the probability that the random variable (X, Y) will assume specific values. As in the one-dimensional case, f_{XY} is nonnegative since it represents a probability. Furthermore, if the density is summed over all possible values of X and Y it must sum to 1. That is, the necessary and sufficient conditions for a function to be a joint density for a two-dimensional discrete random variable are

1. $f_{XY}(x, y) \geq 0$

2. $\sum_{\text{all } x} \sum_{\text{all } y} f_{XY}(x, y) = 1$

The joint density in the discrete case is sometimes expressed in closed form. However, it is more common to present the density in table form.

Example 5.1.1. In an automobile plant, two tasks are performed by robots. The first entails welding two joints; the second, tightening three bolts. Let X denote the number of defective welds and Y the number of improperly tightened bolts produced per car. Since X and Y are each discrete (X, Y) is a two-dimensional discrete random variable. Past data indicates that the joint density for (X, Y) is as shown in Table 5.1. Note that each entry in the table is a number between 0 and 1 and therefore can be interpreted as a probability. Furthermore,

$$\sum_{x=0}^{2} \sum_{y=0}^{3} f_{XY}(x, y) = .840 + .030 + .020 + \cdots + .001 = 1$$

as required. The probability that there will be no errors made by the robots is given by

$$P[X = 0 \text{ and } Y = 0] = f_{XY}(0,0) = .840$$

The probability that there will be exactly one error made is

$$\begin{aligned} P[X = 1 \text{ and } Y = 0] + P[X = 0 \text{ and } Y = 1] &= f_{XY}(1,0) + f_{XY}(0,1) \\ &= .060 + .030 \\ &= .09 \end{aligned}$$

The probability that there will be no improperly tightened bolts is $P[Y = 0]$. Note that this probability, which concerns only the random variable Y, can be obtained

TABLE 5.1

x/y	**0**	**1**	**2**	**3**
0	.840	.030	.020	.010
1	.060	.010	.008	.002
2	.010	.005	.004	.001

y=0

by summing $f_{XY}(x, 0)$ over all values of X. That is,

$$\begin{aligned} P[Y=0] &= \sum_{x=0}^{2} f_{XY}(x,0) \\ &= P[X=0 \text{ and } Y=0] + P[X=1 \text{ and } Y=0] \\ &\quad + P[X=2 \text{ and } Y=0] \\ &= .840 + .060 + .010 = .91 \end{aligned}$$

Given the joint density for a two-dimensional discrete random variable (X, Y), it is easy to derive the individual densities for X and Y. The manner in which this is done is suggested by the method used to answer the last question posed in Example 5.1.1. To find the density for Y alone, we sum the joint density over all values of X; to find the density for X alone we sum over Y. When the joint density is given in table form is customary to report the individual densities for X and Y in the margins of the joint density table. For this reason, the densities for X and Y alone are called *marginal* densities. This idea is formalized in Definition 5.1.2.

Definition 5.1.2 (Discrete marginal densities). Let (X, Y) be a two-dimensional discrete random variable with joint density f_{XY}. The marginal density for X, denoted by f_X, is given by

$$f_X(x) = \sum_{\text{all } y} f_{XY}(x, y)$$

The marginal density for Y, denoted by f_Y, is given by

$$f_Y(y) = \sum_{\text{all } x} f_{XY}(x, y)$$

Example 5.1.2. Table 5.2 gives the joint density for the random variable (X, Y) of Example 5.1.1. It also displays the marginal densities for X, the number of defective welds, and Y, the number of improperly tightened bolts per car. Note that the marginal density for X is obtained by summing across the rows of the table; that for Y is obtained by summing down the columns.

The idea of a two-dimensional continuous random variable and continuous joint density can be developed by extending Definition 4.1.1 to more than one variable.

TABLE 5.2

x/y	**0**	**1**	**2**	**3**	$f_X(x)$
0	.840	.030	.020	.010	.900
1	.060	.010	.008	.002	.080
2	.010	.005	.004	.001	.020
$f_Y(y)$	.910	.045	.032	.013	1.000

Definition 5.1.3 (Continuous joint density). Let X and Y be continuous random variables. The ordered pair (X, Y) is called a two-dimensional continuous random variable. A function f_{XY} such that

1. $f_{XY}(x, y) \geq 0 \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < y < \infty \end{array}$

2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = 1$

3. $\displaystyle P[a \leq X \leq b \text{ and } c \leq Y \leq d] = \int_a^b\int_c^d f_{XY}(x, y)\, dy\, dx$

for a, b, c, d real is called the joint density for (X, Y).

Even though the joint density is defined for all real values x and y, we shall follow the convention of specifying its equation only over those regions for which it may be nonzero. Recall that in the case of a single continuous random variable, probabilities correspond to areas. In the case of a two-dimensional continuous random variable, probabilities correspond to *volumes*. These ideas are illustrated in Example 5.1.3.

Example 5.1.3. In a healthy individual age 20 to 29 years, the calcium level in the blood, X, is usually between 8.5 and 10.5 mg/dl and the cholesterol level, Y, is usually between 120 and 240 mg/dl. Assume that for a healthy individual in this age group the random variable (X, Y) is uniformly distributed over the rectangle whose corners are (8.5, 120), (8.5, 240), (10.5, 120), (10.5, 240). That is, assume that the joint density for (X, Y) is

$$f_{XY}(x, y) = c \qquad \begin{array}{l} 8.5 \leq x \leq 10.5 \\ 120 \leq y \leq 240 \end{array}$$

To be a density, c must be chosen so that

$$\int_{8.5}^{10.5}\int_{120}^{240} c\, dy\, dx = 1$$

That is, c must be chosen so that the volume of the rectangular solid shown in Fig. 5.1(a) is 1. To find c, we can use geometry or complete the indicated integration as shown below.

$$\begin{aligned} \int_{8.5}^{10.5}\int_{120}^{240} c\, dy\, dx &= 1 \\ c\int_{8.5}^{10.5}(240 - 120)\, dx &= 1 \\ 120c(10.5 - 8.5) &= 1 \\ 240c &= 1 \\ c &= 1/240 \end{aligned}$$

Let us now use the joint density to find the probability that an individual's calcium level will lie between 9 and 10 mg/dl while the cholesterol level is between 125 and 140 mg/dl. This probability corresponds to the volume of the solid shown in

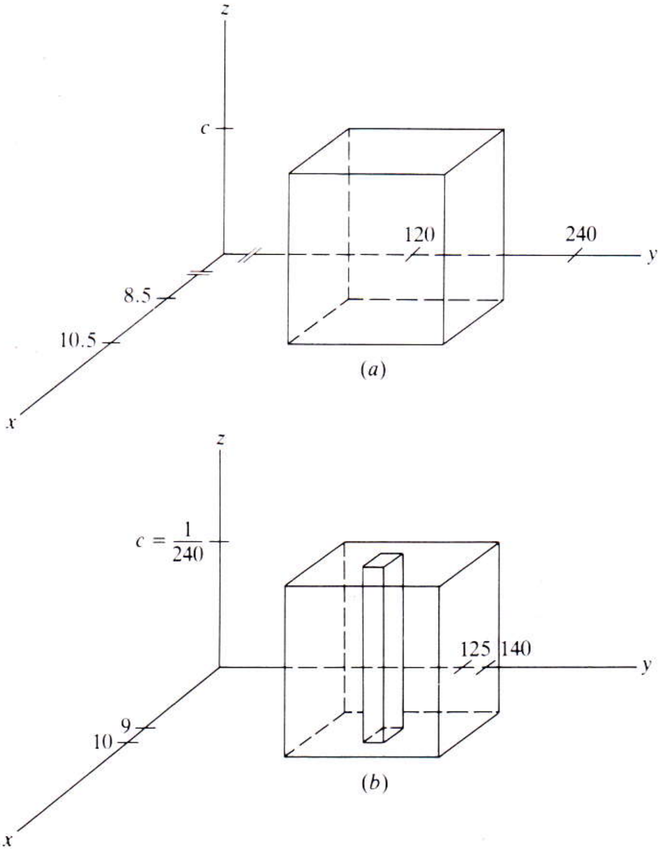


FIGURE 5.1
(*a*) Volume of the solid whose base is a rectangle with corners (8.5, 120), (8.5, 240), (10.5, 120), (10.5, 240) and height c is 1. (*b*) $P[9 \le X \le 10 \text{ and } 125 \le Y \le 140]$ = volume of solid whose base is a rectangle with corners (9, 125), (9, 140), (10, 125), (10, 140) and height $c = 1/240$.

Fig. 5.1(*b*). This probability is

$$
\begin{aligned}
P[9 \le X \le 10 \text{ and } 125 \le Y \le 140] &= \int_9^{10} \int_{125}^{140} 1/240 \, dy \, dx \\
&= 1/240 \int_9^{10} (140 - 125) \, dx \\
&= 15/240
\end{aligned}
$$

To define "marginal" densities in the continuous case, we replace summation by integration. This yields the following definition.

Definition 5.1.4 (Continuous marginal densities). Let (X, Y) be a two-dimensional continuous random variable with joint density f_{XY}. The marginal density for X, denoted by f_X, is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy$$

The marginal density for Y, denoted by f_Y, is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx$$

We illustrate the idea of marginal densities in Examples 5.1.4 and 5.1.5.

Example 5.1.4. Let X denote an individual's blood calcium level and Y his or her blood cholesterol level. The joint density for (X, Y) is

$$f_{XY}(x, y) = 1/240 \qquad 8.5 \le x \le 10.5$$
$$120 \le y \le 240$$

The marginal densities for X and Y are

$$f_X(x) = \int_{120}^{240} 1/240\, dy = 1/2 \qquad 8.5 \le x \le 10.5$$

$$f_Y(y) = \int_{8.5}^{10.5} 1/240\, dx = 2/240 \qquad 120 \le y \le 240$$

To find the probability that a healthy individual has a cholesterol level between 150 and 200, we can use either the joint density or the marginal density for Y. That is,

$$P[150 \le Y \le 200] = \int_{8.5}^{10.5}\int_{150}^{200} 1/240\, dy\, dx = 100/240$$

or

$$P[150 \le Y \le 200] = \int_{150}^{200} 2/240\, dy = 100/240$$

Note that both X and Y are uniformly distributed.

Example 5.1.5. In studying the behavior of air support roofs the random variables X, the inside barometric pressure (in inches of mercury) and Y, the outside pressure, are considered. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = c/x \qquad 27 \le y \le x \le 33$$
$$c = 1/(6 - 27 \ln 33/27) \doteq 1.72$$

The region in the plane over which this joint density is defined is shown in Fig. 5.2(a). The marginal densities for X and Y are given by

$$f_X(x) = \int_{27}^{x} c/x\, dy = (c/x)\, y\Big|_{27}^{x} = c(1 - 27/x) \qquad 27 \le x \le 33$$

$$f_Y(y) = \int_{y}^{33} c/x\, dx = c(\ln 33 - \ln y) \qquad 27 \le y \le 33$$

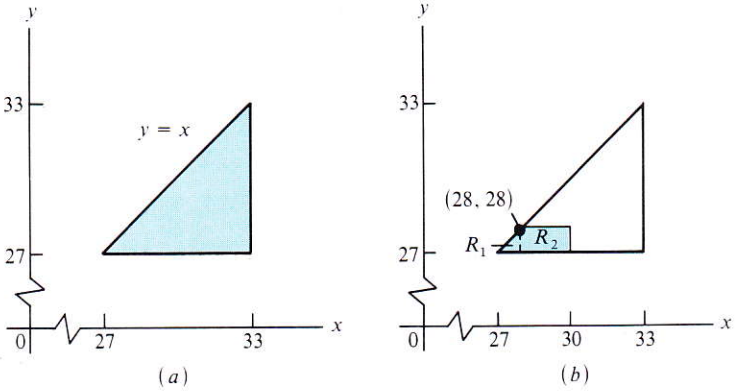


FIGURE 5.2
(a) The joint density $f(x, y) = c/x$ is defined over the triangular region bounded by $y = 27$, $y = x$, and $x = 33$.

(b)

$$P[X \leq 30 \text{ and } Y \leq 28] = \iint_{R_1} c/x\,dy\,dx + \iint_{R_2} c/x\,dy\,dx$$

$$= \int_{27}^{28}\int_{27}^{x} c/x\,dy\,dx + \int_{28}^{30}\int_{27}^{28} c/x\,dy\,dx$$

or

$$P[X \leq 30 \text{ and } Y \leq 28] = \int_{27}^{28}\int_{y}^{30} c/x\,dx\,dy.$$

Let us find the probability that the inside pressure is at most 30 and the outside pressure is at most 28. That is, let us find $P[X \leq 30 \text{ and } Y \leq 28]$. The region over which the joint density is to be integrated is shown in Fig. 5.2(b). Integration can be done with respect to y and then x or vice versa. In the former case, the problem must be split into two pieces since the boundaries for y change at the point (28, 28). In the latter case, integration can be accomplished more easily. The integrals required in the two cases are

Case I

$$P[X \leq 30 \text{ and } Y \leq 28] = \int_{27}^{28}\int_{27}^{x} c/x\,dy\,dx + \int_{28}^{30}\int_{27}^{28} c/x\,dy\,dx$$

Case II

$$P[X \leq 30 \text{ and } Y \leq 28] = \int_{27}^{28}\int_{y}^{30} c/x\,dx\,dy$$

Since case II requires less effort, we find $P[X \leq 30 \text{ and } Y \leq 28]$ as follows:

$$\begin{aligned}
P[X \leq 30 \text{ and } Y \leq 28] &= \int_{27}^{28}\int_{y}^{30} c/x \, dx \, dy \\
&= c\int_{27}^{28}[\ln 30 - \ln y] \, dy \\
&= c\left[y \ln 30\Big|_{27}^{28} - \int_{27}^{28} \ln y \, dy\right] \\
&= c\left[\ln 30 - (y \ln y - y)\Big|_{27}^{28}\right] \\
&= c[\ln 30 - 28 \ln 28 + 27 \ln 27 + 1] \\
&\doteq c(.09) = 1.72(.09) = .15
\end{aligned}$$

It is left as an exercise to show that the same result is obtained via case I. (See Exercise 6.)

Independence

There is one other point to be made in this section. Recall that two events are independent if knowledge that one has occurred gives us no clue as to the likelihood that the other will occur. Suppose that X and Y are discrete random variables such that knowledge of the value assumed by one gives us no clue as to the value assumed by the other. We would like to think of these random variables as being "independent" and would like a mathematical characterization of this property. The characterization is suggested by the following argument. Let X and Y be discrete. Let A_1 denote the event that $X = x$ and let A_2 denote the event that $Y = y$. If X and Y are independent in the intuitive sense, then A_1 and A_2 are independent events. By Definition 2.3.1

$$P[A_1 \cap A_2] = P[A_1]P[A_2]$$

Substituting, it is seen that

$$P[X = x \text{ and } Y = y] = P[X = x]P[Y = y]$$

or

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

It seems that, at least in the discrete case, independence implies that the *joint density can be expressed as the product of the marginal densities*. This idea provides the basis for the definition of the term "independent random variables" in both the discrete and continuous cases.

Definition 5.1.5 (Independent random variables). Let X and Y be random variables with joint density f_{XY} and marginal densities f_X and f_Y respectively. X and Y are independent if and only if

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

for all x and y.

Example 5.1.6

(*a*) The random variables X, the number of defective welds, and Y, the number of improperly tightened bolts per car of Examples 5.1.1 and 5.1.2, are not independent. To verify this, note that from Table 5.2

$$f_{XY}(0,0) = .84 \neq .9(.91) = .819 = f_X(0)f_Y(0)$$

(*b*) The random variables X, an individual's blood calcium level, and Y, his or her blood cholesterol level as described in Examples 5.1.3 and 5.1.4, are independent. To verify this, note that

$$f_{XY}(x,y) = 1/240 = 1/2 \cdot 2/240 = f_X(x)f_Y(y)$$

An important point should be made here. The assumption that (X, Y) is uniformly distributed leads to the conclusion that X and Y are independent. If this conclusion is *medically unsound*, then another more realistic density should be sought to describe the behavior of the two-dimensional random variable (X, Y).

(*c*) The random variables X and Y, the inside and outside pressure respectively on an air support roof of Example 5.1.5, are not independent. This is seen by noting that

$$f_{XY}(x,y) = c/x \neq c(1 - 27/x)c(\ln 33 - \ln y) = f_X(x)f_Y(y)$$

The assumption of nonindependence here is realistic from a physical point of view.

The exercises for Sec. 5.1 provide some practice in dealing with these theoretical ideas. You will see their relationship to data analysis in chapters to come.

5.2 EXPECTATION AND COVARIANCE

In this section we introduce the idea of *expectation* in the case of a two-dimensional random variable. We also study a specific expectation, called the *covariance*, that is useful in describing the behavior of one variable relative to another.

We begin by extending Definitions 3.3.1 and 4.2.1 to the two-dimensional case.

Definition 5.2.1 (Expected value). Let (X, Y) be a two-dimensional random variable with joint density f_{XY}. Let $H(X, Y)$ be a random variable. The expected value of $H(X, Y)$, denoted by $E[H(X, Y)]$ is given by

1. $E[H(X,Y)] = \sum_{\text{all } x} \sum_{\text{all } y} H(x,y) f_{XY}(x,y)$

provided $\sum_{\text{all } x} \sum_{\text{all } y} |H(x,y)| f_{XY}(x,y)$ exists for (X, Y) discrete;

2. $E[H(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x,y) f_{XY}(x,y)\, dy\, dx$

provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |H(x,y)| f_{XY}(x,y)\, dy\, dx$ exists for (X, Y) continuous.

TABLE 5.3

x/y	0	1	2	3	$f_X(x)$
0	.840	.030	.020	.010	.900
1	.060	.010	.008	.002	.080
2	.010	.005	.004	.001	.020
$f_Y(y)$	.910	.045	.032	.013	1.000

Examples 5.2.1 and 5.2.2 illustrate the use of this definition.

Example 5.2.1. The joint density for the random variable (X, Y) of Example 5.1.1 is given in Table 5.3. X denotes the number of defective welds and Y the number of improperly tightened bolts produced per car by assembly line robots. Let us use Definition 5.2.1 to find $E[X]$, $E[Y]$, $E[X+Y]$, and $E[XY]$.

$$\begin{aligned}
E[X] &= \sum_{x=0}^{2} \sum_{y=0}^{3} x f_{XY}(x, y) \\
&= 0(.840) + 0(.030) + 0(.020) + 0(.010) + 1(.060) + \cdots + 2(.001) \\
&= .12 \\
E[Y] &= \sum_{x=0}^{2} \sum_{y=0}^{3} y f_{XY}(x, y) \\
&= 0(.840) + 1(.030) + 2(.020) + 3(.010) + 0(.060) + \cdots + 3(.001) \\
&= .148 \\
E[X+Y] &= \sum_{x=0}^{2} \sum_{y=0}^{3} (x+y) f_{XY}(x, y) \\
&= (0+0)(.840) + (0+1)(.030) + (0+2)(.020) \\
&\quad + \cdots + (2+3)(.001) \\
&= .268 \\
E[XY] &= \sum_{x=0}^{2} \sum_{y=0}^{3} xy f_{XY}(x, y) \\
&= (0 \cdot 0)(.840) + (0 \cdot 1)(.030) + (0 \cdot 2)(.020) + \cdots + (2 \cdot 3)(.001) \\
&= .064
\end{aligned}$$

There are two points to be made. First, both $E[X]$ and $E[Y]$ were found via the joint density and Definition 5.2.1. These expectations could have been found just as easily from the marginal densities and Definition 3.3.1. (See Exercise 18). Second, note that $E[X+Y] = E[X] + E[Y]$. This result is consistent with the rules of expectation given in Theorem 3.3.1.

Example 5.2.2. The joint density for the random variable (X, Y) where X denotes the calcium level and Y the cholesterol level in the blood of a healthy individual is given by

$$f_{XY}(x, y) = 1/240 \qquad 8.5 \le x \le 10.5$$
$$120 \le y \le 240$$

For these variables,

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x, y)\, dy\, dx \\
&= \int_{8.5}^{10.5}\int_{120}^{240} x(1/240)\, dy\, dx \\
&= \int_{8.5}^{10.5} (1/2)\, x\, dx = x^2/4\Big|_{8.5}^{10.5} = 9.5 \text{ mg/dl} \\
E[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x, y)\, dy\, dx \\
&= \int_{8.5}^{10.5}\int_{120}^{240} y(1/240)\, dy\, dx \\
&= 1/240 \int_{8.5}^{10.5} y^2/2\Big|_{120}^{240}\, dx \\
&= 1/240 \int_{8.5}^{10.5} 21{,}600\, dx = 180 \text{ mg/dl} \\
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)\, dy\, dx \\
&= \int_{8.5}^{10.5}\int_{120}^{240} xy(1/240)\, dy\, dx \\
&= 1/240 \int_{8.5}^{10.5} xy^2/2\Big|_{120}^{240}\, dx \\
&= 1/240 \int_{8.5}^{10.5} 21{,}600x\, dx \\
&= (21{,}600/240)(x^2/2)\Big|_{8.5}^{10.5} = 1710
\end{aligned}$$

Occasionally, the expected value of a function of X and Y is of interest in its own right. For instance, in Example 5.2.1, $E[X + Y]$ gives the theoretical average number of errors made by the robots overall. However, we shall be concerned primarily with those expectations that are needed to compute the covariance between X and Y. This term is defined as follows.

Definition 5.2.2 (Covariance). Let X and Y be random variables with means μ_X and μ_Y respectively. The covariance between X and Y, denoted by $\operatorname{Cov}(X, Y)$ or

σ_{XY} is given by

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Note that if small values of X tend to be associated with small values of Y and large values of X with large values of Y, then $X - \mu_X$ and $Y - \mu_Y$ will usually have the same algebraic signs. This implies that $(X - \mu_X)(Y - \mu_Y)$ will be positive, yielding a positive covariance. If the reverse is true and small values of X tend to be associated with large values of Y and vice versa, then $X - \mu_X$ and $Y - \mu_Y$ will usually have opposite algebraic signs. This results in a negative value for $(X - \mu_X)(Y - \mu_Y)$, yielding a negative covariance. In this sense covariance is an indication of how X and Y vary relative to one another.

Covariance is seldom computed from Definition 5.2.2. Rather, we apply the following computational formula whose derivation is left as an exercise. (See Exercise 24.)

Theorem 5.2.1 (Computational formula for covariance)

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

We illustrate the use of Theorem 5.2.1 by finding the covariance for the random variables of Examples 5.2.1 and 5.2.2.

Example 5.2.3

(*a*) The covariance between X, the number of defective welds, and Y, the number of improperly tightened bolts of Example 5.2.1, is given by

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= .064 - (.12)(.148) = .046\end{aligned}$$

Since $\mathrm{Cov}(X, Y) > 0$, there is a tendency for large values of X to be associated with large values of Y and vice versa. That is, a car with an above average number of defective welds tends also to have an above average number of improperly tightened bolts and vice versa.

(*b*) The covariance between X, an individual's blood calcium level, and Y, his or her blood cholesterol level, has covariance given by

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= 1710 - (9.5)(180) = 0\end{aligned}$$

A covariance of 0 implies that knowledge that X assumes a value above its mean gives us no indication as the value of Y relative to its mean.

The fact that the covariance between X and Y is 0 in Example 5.2.2 is not a coincidence. It is, of course, due to the fact that $E[XY] = E[X]E[Y]$. It can be shown that this property will hold whenever the random variables X and Y are independent, as they are in Example 5.2.2. This important result is formalized in the following theorem.

Theorem 5.2.2. Let (X, Y) be a two-dimensional random variable with joint density f_{XY}. If X and Y are independent then

$$E[XY] = E[X]E[Y]$$

Proof. We shall prove this theorem in the continuous case. The proof in the discrete case is similar. Assume that (X, Y) has joint density f_{XY} and that X and Y are independent. Let f_X and f_Y denote the marginal densities for X and Y respectively. By Definition 5.2.1

$$\begin{aligned}
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x, y)\, dy\, dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)\, dy\, dx \qquad (X \text{ and } Y \text{ are independent}) \\
&= \int_{-\infty}^{\infty} xf_X(x)\int_{-\infty}^{\infty} yf_Y(y)\, dy\, dx \\
&= \int_{-\infty}^{\infty} xf_X(x)\, E[Y]\, dx \\
&= E[Y]\int_{-\infty}^{\infty} xf_X(x)\, dx = E[Y]E[X]
\end{aligned}$$

An immediate consequence of this theorem is the result that we have already noted and observed relative to Example 5.2.2. In particular, *if X and Y are independent, then* Cov$(X, Y) = 0$. Unfortunately, the converse of this statement is not true. That is, we *cannot* conclude that a zero covariance implies independence. The next example verifies this contention.

Example 5.2.4. The joint density for (X, Y) is given in Table 5.4, from which we see that $E[X] = 5/2$, $E[Y] = 0$ and $E[XY] = 0$ yielding a covariance of 0. It is also easy to see that X and Y are *not* independent. The value assumed by Y does have an effect on that assumed by X. In fact, $X = Y^2$. The value of Y completely determines the value of X!

Covariance gives us only a very rough idea of the relationship between X and Y. We are concerned only with its algebraic sign and not its magnitude. However, covariance is used to define another measure of the relationship between X and Y which is easier to interpret. This measure, called the *correlation*, is discussed in the next section.

TABLE 5.4

x/y	-2	-1	1	2	$f_X(x)$
1	0	1/4	1/4	0	1/2
4	1/4	0	0	1/4	1/2
$f_Y(y)$	1/4	1/4	1/4	1/4	1

5.3 CORRELATION

Recall that the covariance between X and Y gives only a rough indication of any association that may exist between X and Y. No attempt is made to describe the type or strength of the association. Often it is of interest to know whether or not two random variables are *linearly* related. One measure used to determine this is the Pearson coefficient of correlation, ρ. In this section we define this theoretical measure of linearity; in Chap. 11 we shall discuss how to estimate its value from a data set.

Definition 5.3.1 (Pearson coefficient of correlation). Let X and Y be random variables with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2, respectively. The correlation, ρ_{XY}, between X and Y is given by

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sqrt{(\operatorname{Var} X)(\operatorname{Var} Y)}}$$

Since we already know how to calculate each of the terms appearing in the above definition, calculating ρ_{XY} (or ρ) from the joint density for (X, Y) is easy. The question is: "How do we interpret ρ once we know its numerical value?" To interpret ρ, we must know its range of possible values. The next theorem shows that, unlike the covariance which can assume any real value, the correlation coefficient is bounded.

Theorem 5.3.1. The correlation coefficient ρ_{XY} for any two random variables X and Y lies between -1 and 1 inclusive.

Proof. Let Z and W denote random variables such that $E[Z^2] \neq 0$ and $E[W^2] \neq 0$. Let a denote any real number. Note that since the random variable $(aW - Z)^2 \geq 0$, its mean is nonnegative. That is,

$$E\left[(aW - Z)^2\right] = a^2E[W^2] - 2aE[WZ] + E[Z^2] \geq 0$$

Let $a = E[WZ]/E[W^2]$. Substituting, we can conclude that

$$\frac{(E[WZ])^2}{(E[W^2])^2}E[W^2] - 2\frac{E[WZ]}{E[W^2]}E[WZ] + E[Z^2] \geq 0$$

or

$$-\frac{(E[WZ])^2}{E[W^2]} + E[Z^2] \geq 0$$

This implies that

$$\frac{(E[WZ])^2}{E[W^2]E[Z^2]} \leq 1$$

Now let $W = X - \mu_X$ and $Z = Y - \mu_Y$. Substituting into the above inequality, we can conclude that

$$\frac{\left(E[(X-\mu_X)(Y-\mu_Y)]\right)^2}{\left(E\left[(X-\mu_X)^2\right]\right)\left(E\left[(Y-\mu_Y)^2\right]\right)} = \rho_{XY}^2 \le 1$$

Solving for ρ_{XY}, we see that $|\rho_{XY}| \le 1$ or that $-1 \le \rho_{XY} \le 1$ as claimed.

The next theorem indicates how ρ measures linearity. The point of the theorem is twofold. First, if there is a linear relationship between X and Y, then this fact is reflected in a correlation coefficient of 1 or -1. Second, if $\rho = 1$ or -1, then a linear relationship exists between X and Y. The formal statement of this result is given in Theorem 5.3.2.

Theorem 5.3.2. Let X and Y be random variables with correlation coefficient ρ_{XY}. Then $|\rho_{XY}| = 1$ if and only if $Y = \beta_0 + \beta_1 X$ for some real numbers β_0 and $\beta_1 \ne 0$.

Proof. We shall show that if $|\rho_{XY}| = 1$ then X and Y are linearly related. The proof of the converse is straightforward and is outlined as an exercise (Exercise 35). Assume that $|\rho_{XY}| = 1$. We can reverse the steps given in the proof of Theorem 5.3.1, replacing inequality with equality at each step to conclude that $E[(aW - Z)^2] = 0$. For the mean of a nonnegative random variable to be 0, the variable must equal 0 with probability 1. That is, $P[(aW - Z)^2 = 0] = 1$. This, in turn, implies that $P[aW - Z = 0] = 1$ or that $P[aW = Z] = 1$. Let $W = X - \mu_X$ and $Z = Y - \mu_Y$ to conclude that $P[aX - a\mu_X = Y - \mu_X] = 1$. Rewriting this expression, we can conclude that $P[Y = \mu_Y - a\mu_X + aX] = 1$. That is, $P[Y = \beta_0 + \beta_1 X] = 1$ where $\beta_0 = \mu_Y - a\mu_X$ and $\beta_1 = a$. This means that points not on the line $Y = \beta_0 + \beta_1 X$ occur with 0 probability and the proof is complete.

If $\rho = 1$, then we say that X and Y have *perfect positive* correlation. Perfect position correlation implies that $Y = \beta_0 + \beta_1 X$ where $\beta_1 > 0$. This in turn implies that small values of X are associated with small values of Y, and large values of X with large values of Y. Perfect negative correlation implies that $Y = \beta_0 + \beta_1 X$ where $\beta_1 < 0$. Practically speaking, this means that small values of X are associated with large values of Y and vice versa. Unfortunately, random variables seldom assume the easily interpretable values of 1 or -1. However, values of ρ near 1 or -1 do occur and indicate a linear trend. That is, they indicate that, even though no single straight line passes through the points of positive probability, there is a straight line passing through the graph with the property that most of the probability is associated with points lying on or near this straight line. It is equally important to realize what Theorem 5.3.2 is not saying. If $\rho = 0$, we say that X and Y are uncorrelated, but we are *not* saying that they are unrelated. We are saying that if a relationship exists, then it is *not linear*. These ideas are illustrated in Fig. 5.3.

Example 5.3.1. To find the correlation between X, the number of defective welds, and Y, the number of improperly tightened bolts produced per car by assembly line

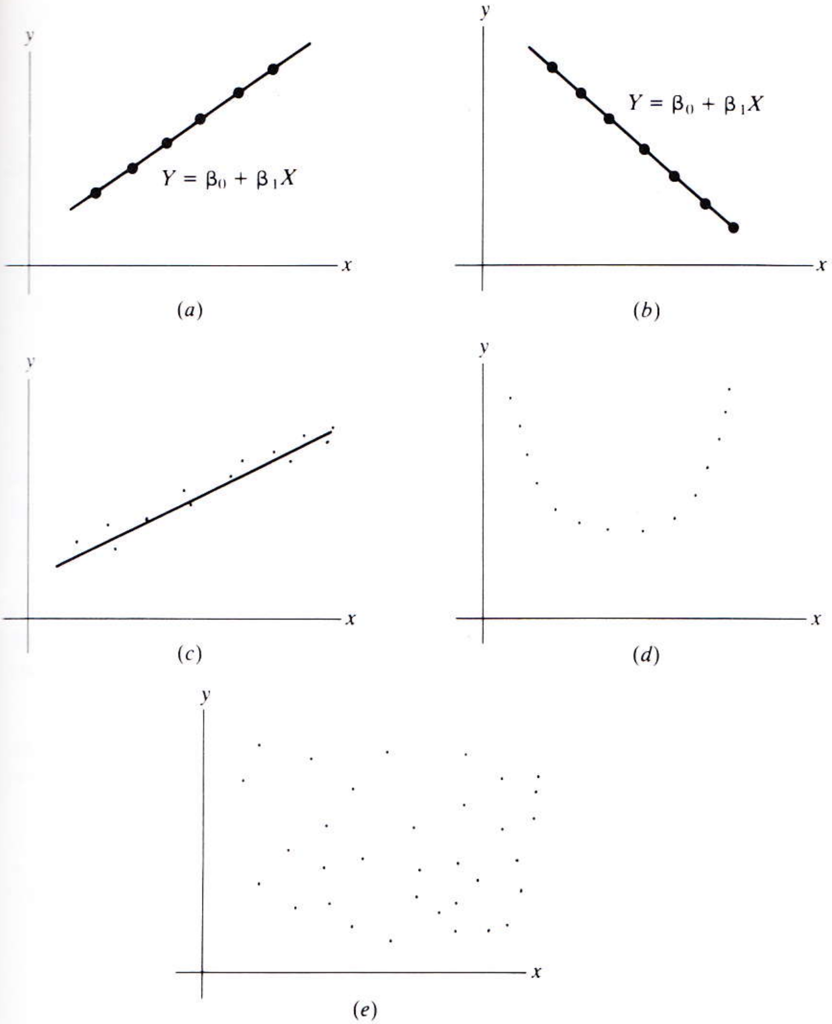


FIGURE 5.3
(a) Perfect positive correlation: $\rho = 1$, $\beta > 0$, all points lie on a straight line with positive slope. (b) Perfect negative correlation: $\rho = -1$, $\beta < 0$, all points lie on a straight line with negative slope. (c) ρ near 1, points exhibit a linear trend. (d) Uncorrelated: $\rho = 0$, points indicate a relationship between X and Y, but the relationship is not linear. (e) Uncorrelated: $\rho = 0$, points are randomly scattered.

robots, we use Table 5.3 to compute $E[X^2]$ and $E[Y^2]$. For these variables

$$E[X^2] = 0^2(.90) + 1^2(.08) + 2^2(.02) = .16$$

$$E[Y^2] = 0^2(.910) + 1^2(.045) + 2^2(.032) + 3^2(.013) = .29$$

In Example 5.2.1, we found that $E[X] = .12$ and $E[Y] = .148$. Therefore

$$\operatorname{Var} X = E[X^2] - (E[X])^2 = .16 - (.12)^2 \doteq .146$$

$$\operatorname{Var} Y = E[Y^2] - (E[Y])^2 = .29 - (.148)^2 \doteq .268$$

In Example 5.2.3, we found that $\mathrm{Cov}(X, Y) = .046$. By Definition 5.3.1

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}\, X\, \mathrm{Var}\, Y}} = \frac{.046}{\sqrt{(.146)(.268)}} \doteq .23$$

Since this value does not appear to lie close to 1, we would not expect the observed values of X and Y to exhibit a strong linear trend.

Exercise 36 points out the relationship between correlation and independence.

5.4 CONDITIONAL DENSITIES AND REGRESSION

In this section we consider two topics that are closely related. These are *conditional densities* and *regression*. To see what is to be done, let us reconsider Example 5.1.5.

Example 5.4.1. In Example 5.1.5 we considered the random variable (X, Y) where X is the inside and Y the outside barometric pressure on an air support roof. Suppose we are interested in studying the inside pressure when the outside pressure is fixed at $y = 30$. There are three important points to understand.

1. The inside pressure will vary even though the outside pressure is constant. Therefore, it makes sense to talk about "the random variable X given that $y = 30$." We shall denote this new random variable by $X|y = 30$.
2. Since $X|y = 30$ is a random variable in its own right, it has a probability distribution. Therefore, it makes sense to ask: "What is the density for $X|y = 30$?" We shall call this density the "conditional density for X given that $y = 30$" and shall denote it by $f_{X|y=30}$.
3. Since the inside pressure varies even though the outside pressure is constant, it makes sense to ask: "What is the mean or average pressure on the inside of the roof when the outside pressure is 30?" That is, we can ask: "What is the mean value for the random variable $X|y = 30$?" This mean value is denoted by $E[X|y = 30]$ or $\mu_{X|y=30}$.

In general, the conditional density for X given $Y = y$, denoted by $f_{X|y}$, is a function that allows us to find the probability that X assumes specific values based on knowledge of the value assumed by the random variable Y. To see how to define $f_{X|y}$ let us assume that (X, Y) is discrete with joint density f_{XY} and marginal densities f_X and f_Y. Let A_1 denote the event that $X = x$ and A_2 denote the event that $Y = y$. From Definition 2.2.1

$$P[A_1|A_2] = \frac{P[A_1 \cap A_2]}{P[A_2]}$$

Substituting, we see that

$$P[X = x|Y = y] = \frac{P[X = x \text{ and } Y = y]}{P[Y = y]} = \frac{f_{XY}(x, y)}{f_Y(y)}$$

In the discrete case, the conditional density for X given $Y = y$, is the ratio of the joint density for (X, Y) to the marginal density for Y. This observation provides the motivation for the definition of the term "conditional density" in both the discrete and continuous cases. In the formal definition, note that the roles of X and Y can be reversed.

Definition 5.4.1 (Conditional density). Let (X, Y) be a two-dimensional random variable with joint density f_{XY} and marginal densities f_X and f_Y. Then

1. The conditional density for X given $Y = y$, denoted by $f_{X|y}$, is given by

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)} \qquad f_Y(y) > 0$$

2. The conditional density for Y given $X = x$, denoted by $f_{Y|x}$, is given by

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \qquad f_X(x) > 0$$

The use of this definition is illustrated in Example 5.4.2.

Example 5.4.2. The joint density for the random variable (X, Y), where X is the inside and Y the outside pressure on an air support roof is given by

$$f_{XY}(x, y) = c/x \qquad 27 \le y \le x \le 33$$

$$c = 1/(6 - 27 \ln 33/27)$$

From Example 5.1.5, the marginal densities for X and Y are

$$f_X(x) = c(1 - 27/x) \qquad 27 \le x \le 33$$

and

$$f_Y(y) = c(\ln 33 - \ln y) \qquad 27 \le y \le 33$$

The conditional density for X given $Y = y$ is

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$$= \frac{c/x}{c(\ln 33 - \ln y)} = \frac{1}{x(\ln 33 - \ln y)} \qquad y \le x \le 33$$

To find the probability that the inside pressure exceeds 32 given that the outside pressure is 30, we let $y = 30$ in the above expression. We then integrate the conditional density over values of X that exceed 32. That is,

$$P[X > 32 | y = 30] = \int_{32}^{33} \frac{1}{x(\ln 33 - \ln 30)}\, dx$$

$$= \frac{\ln x}{\ln 33 - \ln 30}\Bigg|_{32}^{33}$$

$$= \frac{\ln 33 - \ln 32}{\ln 33 - \ln 30} \doteq .32$$

To find the expected or mean value of X given $y = 30$ we apply Definition 4.2.1 to the random variable $X | y = 30$. That is,

$$E[X | y = 30] = \mu_{X|y=30} = \int_{-\infty}^{\infty} x f_{X|y=30}\, dx$$

$$= \int_{30}^{33} x \frac{1}{x(\ln 33 - \ln 30)}\, dx$$

$$= \int_{30}^{33} \frac{1}{\ln 33 - \ln 30}\, dx$$

$$= \frac{3}{\ln 33 - \ln 30} \doteq 31.48$$

When the outside pressure on the roof is 30, the average value of the inside pressure is 31.48 inches of mercury.

In the previous example, note that we did not find the mean for X. We found the mean for X when $y = 30$. The mean value obtained depended on the value chosen for Y. In general the mean of X given $Y = y$ or $\mu_{X|y}$ is a *function of* y. When this function is graphed we obtain what is called the "curve of regression of X on Y." This term is defined formally in Definition 5.4.2. Note that, once again, the roles of X and Y can be reversed.

Definition 5.4.2 (Curve of regression). Let (X, Y) be a two-dimensional random variable.

1. The graph of the mean value of X given $Y = y$, denoted by $\mu_{X|y}$, is called the curve of regression of X on Y.
2. The graph of the mean value of Y given $X = x$, denoted by $\mu_{Y|x}$, is called the curve of regression of Y on X.

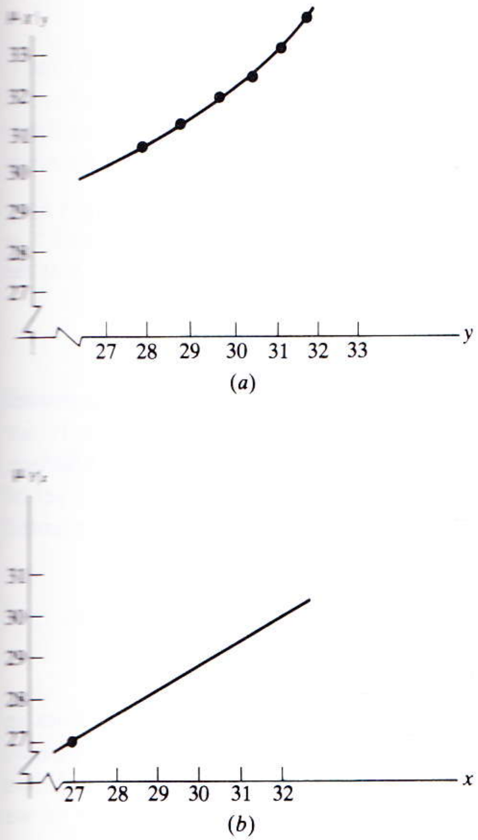
We illustrate the use of this definition by finding the curve of regression of X on Y and the curve of regression of Y on X for the random variable (X, Y) of Example 5.4.2.

Example 5.4.3. The conditional density for X given $Y = y$ where X is the inside and Y is the outside pressure on an air support roof is given by

$$f_{X|y}(x) = \frac{1}{x(\ln 33 - \ln y)} \qquad y \le x \le 33$$

The equation for the curve of regression of X on Y is given by

$$\begin{aligned}\mu_{X|y} &= \int_y^{33} x \frac{1}{x(\ln 33 - \ln y)}\, dx \\ &= \int_y^{33} \frac{1}{\ln 33 - \ln y}\, dx \\ &= \frac{33 - y}{\ln 33 - \ln y}\end{aligned}$$



y	$\mu_{X\|y}$
27	29.90
28	30.43
29	30.95
30	31.47
31	31.98
32	32.50
32.5	32.74

FIGURE 5.4
(*a*) A nonlinear curve of regression: $\mu_{X|y} = (33 - y)/(\ln 33 - \ln y)$. (*b*) A linear curve of regression: $\mu_{Y|x} = (1/2)(x + 27)$.

Note that this equation is *nonlinear*. Its graph is not a straight line. A sketch of the graph is found by plotting $\mu_{X|y}$ for selected values of y. The graph is shown in Fig. 5.4(a). The conditional density for Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$
$$= \frac{c/x}{c(1 - 27/x)}$$
$$= \frac{1}{x - 27} \qquad 27 \le y \le x$$

The equation for the curve of regression of Y on X is given by

$$\mu_{Y|x} = \int_{27}^{x} y \frac{1}{x - 27}\, dy$$
$$= \left. \frac{y^2}{2(x - 27)} \right|_{27}^{x}$$
$$= \frac{x^2 - 27^2}{2(x - 27)}$$
$$= (1/2)(x + 27)$$

Note that this equation is *linear*. Its graph is the straight line shown in Fig. 5.4(b). These curves can be used now to find the mean of X for any specified value of Y or vice versa. For example, the average value of Y, the outside pressure, given that the inside pressure is 29 is

$$\mu_{Y|x=29} = (1/2)(x + 27) = (1/2)(56) = 28 \text{ inches of mercury}$$

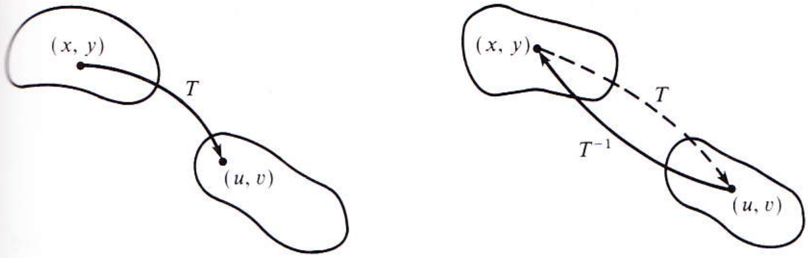
We have introduced only the basic ideas underlying the topic of regression. To find the theoretical regression curves you must *know* the joint density for (X, Y). In practice this density is seldom known with certainty. Thus, in practice, we are forced to approximate these theoretical curves from a data set—a set of observations on the random variable (X, Y). Methods for doing so are presented in Chaps. 11 and 12.

5.5 TRANSFORMATION OF VARIABLES (OPTIONAL)

In Sec. 4.8 we considered the problem of transforming continuous variables in the univariate case. That is, given a continuous random variable X whose density is known, we saw how to find the density for the random variable Y where Y is a function of X. Here we reconsider the problem in the bivariate case. To do so, we must first introduce the notation of Jacobians.

Suppose that we are working in the xy plane and that u and v are variables which are each functions of x and y. That is,

$$u = g_1(x, y) \qquad \text{and} \qquad v = g_2(x, y)$$



$T: \begin{array}{l} u = g_1(x, y) \\ v = g_2(x, y) \end{array}$

(a)

$T^{-1}: \begin{array}{l} x = h_1(u, v) \\ y = h_2(u, v) \end{array}$

(b)

FIGURE 5.5
(a) T maps from the xy plane into the uv plane. (b) T^{-1} maps from the uv plane into the xy plane.

These two equations define a transformation T from some region in the xy plane into the uv plane as pictured in Fig. 5.5(a). Assume that g_1 and g_2 have continuous partial derivatives with respect to x and y. The Jacobian of T is denoted by J_T and is given by the following determinant:

$$J_T = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

Example 5.5.1 illustrates the idea.

Example 5.5.1. Consider the transformation T from the xy plane into the uv plane defined by

$$u = g_1(x, y) = (3y - x)/6$$
$$v = g_2(x, y) = x/3$$

The Jacobian of T is

$$J_T = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} -1/6 & 1/2 \\ 1/3 & 0 \end{vmatrix} = (-1/6)(0) - (1/2)(1/3) = -(1/6)$$

If a transformation T is one-to-one, then it is invertible. Assume that the inverse transformation, T^{-1}, is defined by the equations

$$x = h_1(u, v) \qquad \text{and} \qquad y = h_2(u, v)$$

and that h_1 and h_2 have continuous partial derivatives. [See Fig. 5.5(b).] The

Jacobian of this inverse transformation is given by the determinant

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

This is the sort of Jacobian that will be useful to us in the statistical setting.

Assume that we have two continuous random variables X and Y whose joint density f_{XY} is known. Let U and V be random variables each of which is a function of X and Y. We want to determine the form of f_{UV}, the joint density for (U, V), based on knowledge of the form of f_{XY}. The method for doing so parallels Theorem 4.8.1 and is given in Theorem 5.5.1.

Theorem 5.5.1. Let (X, Y) be continuous with joint density f_{XY}. Let

$$U = g_1(X, Y) \qquad \text{and} \qquad V = g_2(X, Y)$$

where g_1 and g_2 define a one-to-one transformation. Let the inverse transformation be defined by

$$X = h_1(U, V) \qquad \text{and} \qquad V = h_2(U, V)$$

where h_1 and h_2 have continuous first partial derivatives. Then the joint density for (U, V) is given by

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J|$$

where $J \neq 0$ is the Jacobian of the inverse transformation. That is,

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

It is easy to see that Theorem 4.8.1 is a special case of this theorem with f_X corresponding to f_{XY}, $g^{-1}(y)$ playing the role of the inverse transformation, and $|dg^{-1}(y)/dy|$ being equivalent to the absolute value of the Jacobian of the inverse transformation.

Example 5.5.2. Assume that X and Y are independent uniformly distributed random variables over $(0, 2)$ and $(0, 3)$, respectively. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 1/6 \qquad \begin{array}{l} 0 < x < 2 \\ 0 < y < 3 \end{array}$$

Let $U = X - Y$ and $V = X + Y$. What is the joint density for (U, V)? To apply Theorem 5.5.1, we first note that the transformation

$$T: \begin{cases} U = X - Y \\ V = X + Y \end{cases}$$

is a linear transformation from the xy plane into the uv plane. A result from advanced calculus states that a linear transformation from two-dimensional space into two-dimensional space is one-to-one whenever the determinant of its matrix of

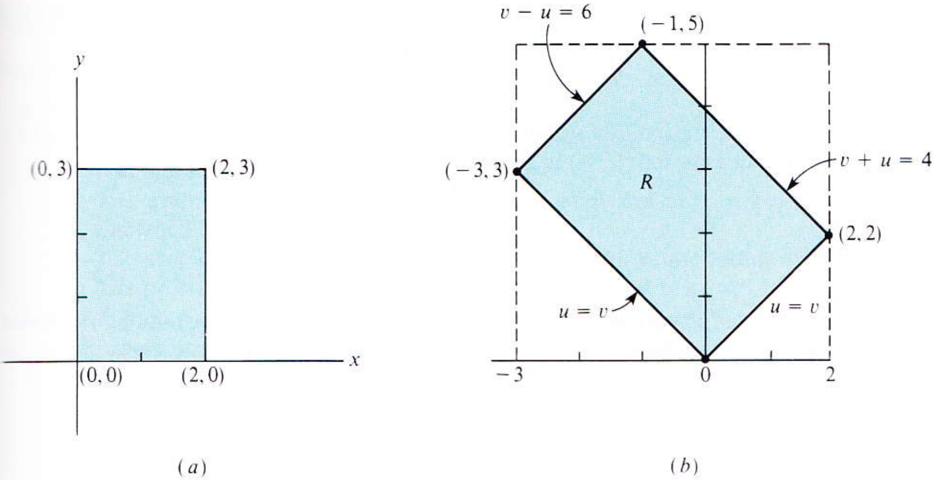


FIGURE 5.6
(a) (X, Y) lies in the rectangle with corners (0, 0), (2, 0), (0, 3), (2, 3). (b) (U, V) lies in the region R.

coefficients is not zero. Here the determinant is

$$\begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = (1)(1) - (1)(-1) = 2$$

so T is invertible. The inverse Transformation is found by solving the above system of equations for X and Y. Here T^{-1} is given by

$$T^{-1}: \begin{cases} X = (U + V)/2 \\ Y = (U - V)/2 \end{cases}$$

The Jacobian of T^{-1} is

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{vmatrix} = (1/2)(1/2) - (1/2)(-1/2) = 1/2$$

By Theorem 5.5.1

$$\begin{aligned} f_{UV}(u, v) &= f_{XY}((v + u)/2, (v - u)/2)|J| \\ &= (1/6)(1/2) = 1/12 \end{aligned}$$

To find the set of values for which $f_{UV} > 0$, we note that since $0 < x < 2$ and $0 < y < 3$, (X, Y) lies in the rectangle shown in Fig. 5.6(a). It is easy to see that $U = X - Y$ must lie between -3 and 2 and that $V = X + Y$ must lie between 0 and 5. Furthermore, U and V must satisfy the inequalities

$$0 < (v + u)/2 < 2$$

$$0 < (v - u)/2 < 3$$

or

$$0 < v + u < 4$$
$$0 < v - u < 6$$

Solving these inequalities simultaneously yields the region R shown in Fig. 5.6(b). Thus the density for (U, V) is given by

$$f_{UV}(u, v) = 1/12 \qquad (u, v) \in R$$

We leave it to you to verify that f_{UV} is, in fact, a valid density.

Other transformation theorems can be derived from Theorem 5.5.1. Some of these are given in Exercises 48, 50, and 51. For a more detailed discussion of this topic, please see [47].

CHAPTER SUMMARY

In this chapter we considered random variables of more than one dimension. Emphasis was on random variables of two dimensions. The joint density was defined by extending the notion of a density for a single variable in a logical way. This function was used to calculate probabilities associated with two-dimensional random variables (X, Y). We saw how to obtain the marginal densities for both X and Y from the joint density. These marginal densities are the usual densities for X or Y when considered alone. The correlation coefficient ρ was introduced as a measure of linearity between X and Y. The notion of independence between X and Y was defined formally and its relationship to ρ was investigated. We saw how to define the conditional densities for X given Y and Y given X from knowledge of the joint density for (X, Y) and the marginal densities for X and Y. The conditional densities were used to find the equations for the curves of regression of Y on X and X on Y. These regression curves are the graphs of the mean value of Y as a function of X or vice versa. We say that these curves may be linear or nonlinear.

Terms with which you should be familiar now are these:

Two-dimensional discrete random variable
Two-dimensional continuous random variable
Discrete joint density
Continuous joint density
Discrete marginal density
Continuous marginal density
Independent random variables
Expected value of $H(X, Y)$
Covariance
Correlation coefficient
Perfect positive correlation
Perfect negative correlation
Uncorrelated
Conditional density
Curve of regression
n-dimensional discrete random variable
n-dimensional continuous random variable
Bivariate normal distribution

EXERCISES

Section 5.1

1. Use Table 5.2 to find each of these probabilities:
   (*a*) The probability that exactly two defective welds and one improperly tightened bolt will be produced by the robots.
   (*b*) The probability that at least one defective weld and at least one improperly tightened bolt will be produced.
   (*c*) The probability that at most one defective weld will be produced.
   (*d*) The probability that at least two improperly tightened bolts will be produced.
2. In conducting an experiment in the laboratory, temperature gauges are to be used at four junction points in the equipment setup. These four gauges are randomly selected from a bin containing seven such gauges. Unknown to the scientist, three of the seven gauges give improper temperature readings. Let X denote the number of defective gauges selected and Y the number of nondefective gauges selected. The joint density for (X, Y) is given in Table 5.5.

TABLE 5.5

x/y	0	1	2	3	4
0	0	0	0	0	1/35
1	0	0	0	12/35	0
2	0	0	18/35	0	0
3	0	4/35	0	0	0

   (*a*) The values given in Table 5.5 can be derived by realizing that the random variable X is hypergeometric. Use the results of Sec. 3.7 to verify the values given in Table 5.5.
   (*b*) Find the marginal densities for both X and Y. What type of random variable is Y?
   (*c*) Intuitively speaking, are X and Y independent? Justify your answer mathematically.
3. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 1/n^2 \qquad x = 1, 2, 3, \ldots, n$$
$$y = 1, 2, 3, \ldots, n$$

   (*a*) Verify that $f_{XY}(x, y)$ satisfies the conditions necessary to be a density.
   (*b*) Find the marginal densities for X and Y.
   (*c*) Are X and Y independent?

*4. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 2/n(n+1) \qquad 1 \le y \le x \le n \qquad n \text{ a positive integer}$$

   (*a*) Verify that $f_{XY}(x, y)$ satisfies the conditions necessary to be a density. *Hint:* The sum of the first n integers is given by $n(n+1)/2$.
   (*b*) Find the marginal densities for X and Y. *Hint:* Draw a picture of the region over which (X, Y) is defined.
   (*c*) Are X and Y independent?

(*d*) Assume that $n = 5$. Use the joint density to find $P[X \le 3 \text{ and } Y \le 2]$. Find $P[X \le 3]$ and $P[Y \le 2]$. *Hint:* Draw a picture of the region over which (X, Y) is defined.

5. The two most common types of errors made by programmers are syntax errors and errors in logic. For a simple language such as BASIC the number of such errors is usually small. Let X denote the number of syntax errors and Y the number of errors in logic made on the first run of a BASIC program. Assume that the joint density for (X, Y) is as shown in Table 5.6.

TABLE 5.6

x/y	**0**	**1**	**2**	**3**
0	.400	.100	.020	.005
1	.300	.040	.010	.004
2	.040	.010	.009	.003
3	.009	.008	.007	.003
4	.008	.007	.005	.002
5	.005	.002	.002	.001

(*a*) Find the probability that a randomly selected program will have neither of these types of errors.

(*b*) Find the probability that a randomly selected program will contain at least one syntax error and at most one error in logic.

(*c*) Find the marginal densities for X and Y.

(*d*) Find the probability that a randomly selected program contains at least two syntax errors.

(*e*) Find the probability that a randomly selected program contains one or two errors in logic.

(*f*) Are X and Y independent?

6. Consider Example 5.1.5. Verify that $P[X \le 30 \text{ and } Y \le 28] \doteq .15$ by integrating the joint density first with respect to y then with respect to x.

7. (*a*) Use the joint density of Example 5.1.5 to find the probability that the inside pressure on the roof will be greater than 30 while the outside pressure is less than 32.

(*b*) Use the marginal density for X to find $P[X \le 28]$.

(*c*) Use the marginal density for Y to find $P[Y > 30]$.

8. Let X denote the temperature (°C) and let Y denote the time in minutes that it takes for the diesel engine on an automobile to get ready to start. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = c(4x + 2y + 1) \qquad 0 \le x \le 40$$
$$0 \le y \le 2$$

(*a*) Find the value of c that makes this a density.

(*b*) Find the probability that on a randomly selected day, the air temperature will exceed 20°C and it will take at least one minute for the car to be ready to start.

(*c*) Find the marginal densities for X and Y.

(*d*) Find the probability that on a randomly selected day it will take at least one minute for the car to be ready to start.

(*e*) Find the probability that on a randomly selected day, the air temperature will exceed 20°C.
(*f*) Are X and Y independent? Explain on a mathematical basis.

9. An engineer is studying early morning traffic patterns at a particular intersection. The observation period begins at 5:30 a.m. Let X denote the time of arrival of the first vehicle from the north-south direction; let Y denote the first arrival time from the east-west direction. Time is measured in fractions of an hour after 5:30 a.m. Assume that the density for (X, Y) is given by

$$f_{XY}(x, y) = 1/x \qquad 0 < y < x < 1$$

(*a*) Verify that this a joint density for a two-dimensional random variable.
(*b*) Find $P[X \leq .5 \text{ and } Y \leq .25]$.
(*c*) Find $P[X > .5 \text{ or } Y > .25]$.
(*d*) Find $P[X \geq .5 \text{ and } Y \geq .5]$.
(*e*) Find the marginal densities for X and Y.
(*f*) Find $P[X \leq .5]$.
(*g*) Find $P[Y \leq .25]$.
(*h*) Are X and Y independent? Explain.

10. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = x^3y^3/16 \qquad 0 \leq x \leq 2, 0 \leq y \leq 2$$

(*a*) Find the marginal densities for X and Y.
(*b*) Are X and Y independent?
(*c*) Find $P[X \leq 1]$.
(*d*) If it is known that $Y = 1$, what is $P[X \leq 1]$? (Don't do any computation to answer this question!)

11. Economic conditions cause fluctuations in the prices of raw commodities as well as in finished products. Let X denote the price paid for a barrel of crude oil by the initial carrier and let Y denote the price paid by the refinery purchasing the product from the carrier. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = c \qquad 20 < x < y < 40$$

(*a*) Find the value of c that makes this a joint density for a two-dimensional random variable.
(*b*) Find the probability that the carrier will pay at least \$25 per barrel and the refinery will pay at most \$30 per barrel for the oil.
(*c*) Find the probability that the price paid by the refinery exceeds that of the carrier by at least \$10 per barrel.
(*d*) Find the marginal densities for X and Y.
(*e*) Find the probability that the price paid by the carrier is at least \$25.
(*f*) Find the probability that the price paid by the refinery is at most \$30.
(*g*) Are X and Y independent? Explain.

*12. (*n-dimensional discrete random variables.*) Random variables of dimension $n > 2$ can be defined and studied by extending the definitions presented in the two-dimensional case in a logical way. For example, an n-tuple $(X_1, X_2, X_3, \ldots, X_n)$ in which each of the random variables $X_1, X_2, X_3, \ldots, X_n$ is a discrete random variable is called an n-dimensional discrete random variable. The density for such a random variable is given by

$$f(x_1, x_2, x_3, \ldots, x_n) = P[X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n]$$

This problem entails the use of a 3-dimensional random variable.

Items coming off an assembly line are classed as being either nondefective, defective but salvageable, or defective and nonsalvageable. The probabilities of observing items in each of these categories are .9, .08, and .02 respectively. The probabilities do not change from trial to trial. Twenty items are randomly selected and classified. Let X_1 denote the number of nondefective items obtained, X_2 the number of defective but salvageable items obtained, and X_3 the number of defective and nonsalvageable items obtained.

(*a*) Find $P[X_1 = 15, X_2 = 3, X_3 = 2]$. *Hint:* Use the formula for the number of permutations of indistinguishable objects, Exercise 17, Chap. 1, to count the number of ways to get this sort of split in a sequence of 20 trials.

(*b*) Find the general formula for the density for (X_1, X_2, X_3).

***13.** (*n-dimensional continuous random variables.*) An n-tuple $(X_1, X_2, X_3, \ldots, X_n)$ where each of the random variables $X_1, X_2, \ldots, X_n$ is continuous is called an n-dimensional continuous random variable. The density for an n-dimensional continuous random variable is defined by extending Definition 5.1.3 in a natural way. State the three properties that identify a function as a density for $(X_1, X_2, X_3, \ldots, X_n)$.

***14.** Let $f(x_1, x_2, x_3) = c(x_1 \cdot x_2 \cdot x_3)$ for $0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1$. Find the value of c that makes this a density for the three-dimensional random variable (X_1, X_2, X_3).

Section 5.2

15. Four temperature gauges are randomly selected from a bin containing three defective and four nondefective gauges. Let X denote the number of defective gauges selected and Y the number of nondefective gauges selected. (See Exercise 2.) The joint density for (X, Y) is given in Table 5.5.

(*a*) From the physical description of the problem, should $\text{Cov}(X, Y)$ be positive or negative?

(*b*) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.

16. Let X denote the number of syntax errors and Y the number of errors in logic made on the first run of a BASIC program. (See Exercise 5.) The joint density for (X, Y) is given in Table 5.6.

(*a*) X and Y are not independent. Does this give any indication of the value of the covariance?

(*b*) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$. Give a rough physical interpretation of the covariance.

(*c*) Find $E[X + Y]$. What is the practical interpretation of this expectation?

17. Consider the random variable (X, Y) of Exercise 3. Without doing any additional computation, find $\text{Cov}(X, Y)$.

18. Use the marginal densities given in Table 5.3 to compute $E[X]$ and $E[Y]$. Compare your results to those obtained in Example 5.2.1.

19. The joint density for (X, Y), where X is the inside and Y is the outside barometric pressure on an air support roof (see Example 5.1.5), is given by

$$f_{XY}(x, y) = c/x \qquad 27 \le y \le x \le 33$$

$$c = 1/(6 - 27 \ln 33/27)$$

(a) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.
(b) Find $E[X - Y]$. What is the practical physical interpretation of this expectation?

20. The joint density for (X, Y), where X is the temperature and Y is the time that it takes for a diesel engine on an automobile to get ready to start (see Exercise 8), is given by

$$f_{XY}(x, y) = (1/6640)(4x + 2y + 1) \qquad 0 \le x \le 40$$
$$0 \le y \le 2$$

(a) From a physical standpoint do you think $\text{Cov}(X, Y)$ should be positive or negative?
(b) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.

21. The joint density for (X, Y), where X is the arrival time of the first vehicle from the north-south direction and Y is the arrival time of the first vehicle from the east-west direction at an intersection (see Exercise 9), is given by

$$f_{XY}(x, y) = 1/x \qquad 0 < y < x < 1$$

Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.

22. Find the covariance between the random variables X and Y of Exercise 10.

23. Let X denote the price paid for a barrel of crude oil by the initial carrier and let Y denote the price paid by the refinery purchasing the oil. (See Exercise 11.) The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 1/200 \qquad 20 < x < y < 40$$

(a) From a physical standpoint should $\text{Cov}(X, Y)$ be positive or negative?
(b) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.
(c) Find $E[Y - X]$. Interpret this expectation in a practical sense.

24. Show that $\text{Cov}(XY) = E[XY] - E[X]E[Y]$. *Hint:* By definition, $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$. Expand this product and apply the rules for expectation (Theorem 3.3.1). Remember that $\mu_X = E[X]$ and $\mu_Y = E[Y]$.

25. Prove that $\text{Var}(X + Y) = \text{Var}\, X + \text{Var}\, Y + 2\,\text{Cov}(X, Y)$. *Hint:* $\text{Var}(X + Y) = \text{E}[(X + Y)^2] - (E[X + Y])^2$. Square these terms and apply the rules for expectation. (Theorem 3.3.1.)

26. Use the result of Exercise 25 to show that if X and Y are independent then $\text{Var}(X + Y) = \text{Var}\, X + \text{Var}\, Y$. This proves the third rule for variance. (Theorem 3.3.3.)

27. Show that if $X = Y$, then $\text{Cov}(X, Y) = \text{Var}\, X = \text{Var}\, Y$.

*28. Let the joint density for (X, Y) be given by

$$f(x, y) = \frac{1}{2(e-1)}\left[\frac{1}{x} + \frac{1}{y}\right] \qquad 1 \le x \le e \qquad 1 \le y \le e$$

(a) Show that $\int_1^e \int_1^e f(x, y)\, dy\, dx = 1$.
(b) Find $E[X]$ and $E[Y]$.
(c) Find $E[XY]$.
(d) Are X and Y independent? Explain, based on your answers to parts (b) and (c) and Theorem 5.2.2.

Section 5.3

29. The joint density for (X, Y), where X denotes the number of defective and Y the number of nondefective temperature gauges selected from a bin containing three defective and four nondefective gauges, is given in Table 5.5. (See Exercise 2.)
  (*a*) From the physical interpretation of the problem, should ρ_{XY} be positive or negative? Should ρ_{XY} be $+1$ or -1? Explain.
  (*b*) Find $E[X^2]$ and $E[Y^2]$. Use the information from Exercise 15 to find ρ_{XY}.

In Exercises 30 to 34, find $E[X^2]$, $E[Y^2]$, Var X, Var Y, and ρ_{XY} for the random variables in the exercises referenced. In each case decide whether or not you would expect the graph of Y versus X to exhibit a strong linear trend.

30. Exercise 16.

31. Exercise 19.

32. Exercise 20.

33. Exercise 21.

34. Exercise 23.

35. Assume that $Y = \beta_0 + \beta_1 X$, $\beta_1 \neq 0$.
  (*a*) Show that $\text{Cov}(X, Y) = \beta_1 \text{Var } X$. *Hint:* $\text{Cov}(X, Y) = E[X(\beta_0 + \beta_1 X)] - E[X] \times E[\beta_0 + \beta_1 X]$. Use the rules for expectation.
  (*b*) Show that $\text{Var } Y = \beta_1^2 \text{Var } X$. *Hint:* Use the rules for variance. (Theorem 3.3.3.)
  (*c*) Find ρ_{XY}.
  (*d*) Argue that $\rho_{XY} = 1$ if β_1, the slope of the line $Y = \beta_0 + \beta_1 X$, is positive and that $\rho_{XY} = -1$ if the slope of this line is negative.

36. Prove that if X and Y are independent, then $\rho_{XY} = 0$. Can we conclude that if X and Y are uncorrelated then they are independent? Explain.

37. Without doing any additional computation, find ρ_{XY} for the random variables of Exercise 3.

38. What is the correlation between the random variables X and Y of Exercise 10?

Section 5.4

39. Consider Example 5.4.3.
  (*a*) What is the expected value of X when $y = 31$?
  (*b*) What is the expected value of Y when $x = 30$?

40. Consider Example 5.1.4.
  (*a*) Find $f_{X|y}$. Note that $f_{X|y} = f_X$. From a physical standpoint, can you explain why these densities are the same?
  (*b*) Find $f_{Y|x}$. Is $f_{Y|x} = f_Y$?
  (*c*) Find the curve of regression of X on Y and the curve of regression of Y on X. Are these curves linear?

***41.** Consider the random variable (X, Y) of Exercise 4.
  (*a*) Find the curve of regression of X on Y. Is the regression linear?
  (*b*) Assume that $n = 10$ and find the mean value of X when $y = 4$.
  (*c*) Find the curve of regression of Y on X. Is the regression linear?
  (*d*) Assume that $n = 10$ and find the mean value of Y when $x = 4$.

42. Consider the random variable (X, Y) of Exercise 9.
   (*a*) Find the curve of regression of X on Y. Is the regression linear?
   (*b*) Find the mean value of X when $y = .5$.
   (*c*) Find the curve of regression of Y on X. Is the regression linear?
   (*d*) Find the mean value of Y when $x = .75$.
43. Consider Exercise 11.
   (*a*) Find the curve of regression of X on Y. Is the regression linear?
   (*b*) Find the mean price paid by the carrier for a barrel of crude oil given that the refinery price is \$30 per barrel.
   (*c*) Find the curve of regression of Y on X. Is the regression linear?
   (*d*) Find the mean price paid by the refinery for a barrel of crude oil given that the carrier paid \$35 per barrel.
44. Note that if $|\rho| = 1$, then $Y = \beta_0 + \beta_1 X$. For fixed values of X, $Y|x = \beta_0 + \beta_1 x$. Argue that $\mu_{Y|x}$ is a linear function of x. That is, argue that if X and Y are perfectly correlated then the curve of regression of Y on X is linear. Is the converse true? Explain.

Section 5.5

45. Consider the linear transformation T defined by

$$T:\ u = 2x + y$$
$$v = x + 3y$$

   (*a*) Is this transformation invertible? If so, find the defining equations for T^{-1}.
   (*b*) Find the Jacobian for T^{-1}.
46. Consider the linear transformation T defined by

$$T:\ u = 3x + 2y$$
$$v = x - y$$

   (*a*) Is this transformation invertible? If so, find the defining equations for T^{-1}.
   (*b*) Find the Jacobian for T^{-1}.
47. Assume that X and Y are independent and uniformly distributed over $(0, 1)$ and $(0, 2)$, respectively. Find the joint density for (U, V) where U and V are as defined in Exercise 45.
48. (*Distribution of one function of two continuous random variables.*) Let X and Y be continuous random variables with joint density f_{XY}. Let $U = X + Y$. Prove that f_U, the density for $X + Y$, is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v)\, dv$$

   Hint: Define a transformation T by

$$u = g_1(x, y) = x + y$$
$$v = g_2(x, y) = y$$

   Follow the procedure given in Theorem 5.5.1 to obtain the joint density for (U, V). Integrate the joint density to obtain the marginal density for U.
49. Let X and Y be independent standard normal random variables. Let $U = X + Y$. Use Exercise 48 to prove that U follows a normal distribution with mean 0 and

variance 2. *Hint:* In integrating over v, complete the square in the exponent and remember that a normal density integrated over the real line is equal to 1.

50. Let X and Y be continuous random variables with joint density f_{XY}. Let $U = XY$. Prove that f_U, the density for XY, is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u/v, v)|1/v|\, dv$$

Hint: Let $u = g_1(x, y) = xy$ and $v = y$ and apply Theorem 5.5.1.

51. Let X and Y be continuous random variables with joint density f_{XY}. Let $U = X/Y$. Prove that f_U, the density for X/Y, is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v)|v|\, dv$$

Hint: Let $u = g_1(x, y) = x/y$ and $v = y$ and apply Theorem 5.5.1.

52. Let X and Y be independent exponentially distributed random variables with parameters β_1 and β_2, respectively.

(a) Find the joint density for (X, Y).

(b) Let $U = X + Y$ and verify that

$$f_U(u) = \int_0^u f_{XY}(u - v, v)\, dv$$

Hint: Remember that $0 < x < \infty$ and that $x = u - v$.

(c) Assume that $\beta_1 = 3$ and $\beta_2 = 1$. Show that

$$f_U(u) = e^{-u/3} - e^{-u/2} \qquad 0 < u < \infty$$

53. Let X and Y be independent uniformly distributed random variables over the intervals (0, 2) and (0, 3), respectively.

(a) Let $U = XY$ and find f_U.

(b) Let $U = X/Y$ and find f_U.

REVIEW EXERCISES

54. An electronic device is designed to switch house lights on and off at random times after it has been activated. Assume that it is designed in such a way that it will be switched on and off exactly once in a one-hour period. Let Y denote the time at which the lights are turned on and X the time at which they are turned off. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = 8xy \qquad 0 < y < x < 1$$

(a) Verify that f_{XY} satisfies the conditions necessary to be a density.

(b) Find $E[XY]$.

(c) Find the probability that the lights will be switched on within 1/2 hour after being activated and then switched off again within 15 minutes.

(d) Find the marginal density for X. Find $E[X]$ and $E[X^2]$.

(e) Find the marginal density for Y. Find $E[Y]$ and $E[Y^2]$.

(*f*) Are X and Y independent?

(*g*) Find the conditional distribution of X given Y.

(*h*) Find the probability that the lights will be switched off within 45 minutes of the system being activated given that they were switched on 10 minutes after the system was activated.

(*i*) Find the curve of regression of X on Y. Is the regression linear?

(*j*) Find the expected time that the lights will be turned off given that they were turned on 10 minutes after the system was activated.

(*k*) Based on the physical description of the problem would you expect ρ to be positive, negative, or 0? Explain. Verify by computing ρ.

55. Verify that

$$f_{XY}(x, y) = xye^{-x}e^{-y} \qquad x > 0 \qquad y > 0$$

satisfies the conditions necessary to be a density for a continuous random variable (X, Y). Find the marginal densities for X and Y. Are X and Y independent? Find ρ_{XY}.

56. Let X denote the number of "do loops" in a Fortran program and Y the number of runs needed for a novice to debug the program. Assume that the joint density for (X,Y) is given in Table 5.7.

TABLE 5.7

x/y	1	2	3	4
0	.059	.100	.050	.001
1	.093	.120	.082	.003
2	.065	.102	.100	.010
3	.050	.075	.070	.020

(*a*) Find the probability that a randomly selected program contains at most one "do loop" and requires at least two runs to debug the program.

(*b*) Find $E[XY]$.

(*c*) Find the marginal densities for X and Y. Use these to find the mean and variance for both X and Y.

(*d*) Find the probability that a randomly selected program requires at least two runs to debug given that it contains exactly one "do loop."

(*e*) Find Cov(X, Y). Find the correlation between X and Y. Based on the observed value of ρ, can you claim that X and Y are not independent? Explain.

57. Vehicles arrive at a highway toll booth at random instances from both the south and north. Assume that they arrive at average rates of five and three per five-minute period respectively. Let X denote the number arriving from the south during a five-minute period and let Y denote the number arriving from the north during this same time. Assume that X and Y are independent.

(*a*) Find the joint density for (X, Y).

(*b*) Find the probability that a total of four vehicles arrives during a five-minute time period.

(*c*) Find the correlation between X and Y.

(*d*) Find the conditional density for X given $Y = y$.

***58.** (*Bivariate normal distribution.*) A random variable (X, Y) is said to have a bivariate normal distribution if its joint density is given by

$$f_{XY}(x, y) = \frac{\exp\left\{\dfrac{1}{-2(1-\rho^2)}\left[\left(\dfrac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\dfrac{x-\mu_X}{\sigma_X}\right)\left(\dfrac{y-\mu_Y}{\sigma_Y}\right) + \left(\dfrac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

where x and y can assume any real value. The parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ denote the respective means and standard deviations for X and Y. The parameter ρ is the correlation coefficient. The name of this distribution comes from the fact that the marginal densities for X and Y are both normal. Show that in the case of a bivariate normal distribution, if $\rho = 0$, then X and Y are independent.

CHAPTER 6

DESCRIPTIVE STATISTICS

Thus far, we have considered random variables from a theoretical point of view. We have studied two functions, the density and the cumulative distribution function, that enable us to predict the behavior of the variable in a probabilistic sense. We have also considered three parameters that characterize or describe a random variable, namely, μ, σ^2, and σ. In practice, the exact distribution of a random variable is seldom known. Rather, we must determine a reasonable form for the density and appropriate values for the distribution parameters from a data set. In this chapter we consider some simple graphical and analytic methods for doing so.

6.1 RANDOM SAMPLING

We begin by considering a typical problem that calls for a statistical solution. Suppose that we wish to study the performance of the lithium batteries used in a particular model of pocket calculator. The purpose of our study is to determine the mean effective life span of these batteries so that we can place a limited warrantee on them in the future. Since this type of battery has not been used in this model before, no one can tell us the distribution of the random variable, X, the life span of a battery. We must attempt to discover its distribution for ourselves. This is inherently a statistical problem. What characteristics identify it as such? Simply these:

1. Associated with the problem is a large group of objects about which inferences are to be made. This group of objects is called the *population*.
2. There is at least one random variable whose behavior is to be studied relative to the population.

3. The population is too large to study in its entirety, or techniques used in the study are destructive in nature. In either case, we must draw conclusions about the population based on observing only a portion or "sample" of objects drawn from the population.

In our example, the population is large and hypothetical in the sense that it consists of all lithium batteries used in this model calculator in the past, present, and future. Since we cannot observe the life span of batteries not yet produced, the population obviously cannot be studied in its entirety! Furthermore, to determine the life span of a battery, it must be used until it fails. That is, the method of study destroys the object being studied. For these reasons, we must devise methods for approximating the characteristics of the life span of a lithium battery based on observing only a sample of these batteries.

To draw inferences about a population using statistical methods, the sample drawn should be "random." To understand what is meant by this term, let us

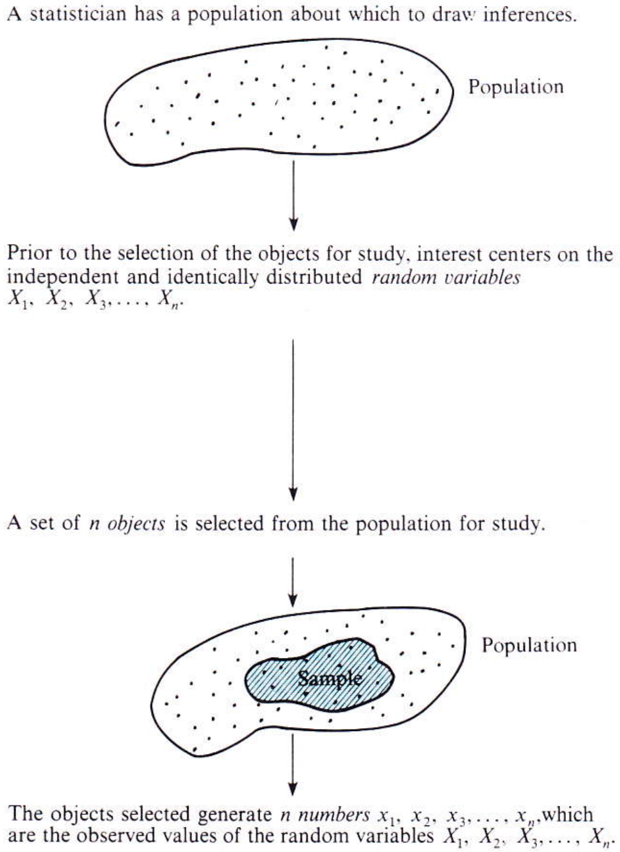


FIGURE 6.1

return to our example. Here we have a large population that consists of all lithium batteries produced for a certain model of pocket calculator. Associated with the population is a random variable X. We do not know the form of its density, nor do we know its mean or variance. We want to select a subset of n batteries from the population "at random." That is, we want to select n batteries for study in such a way that the selection of one battery neither ensures nor precludes the selection of any other. In this way the selection of one battery is independent of the selection of any other. This collection of objects can be thought of as a "random sample."

Note that, prior to the actual selection of the batteries to be studied, X_i ($i = 1, 2, 3, \ldots, n$), the life span of the ith battery selected is a random variable. It has the same distribution as X, the life span of batteries in the population. Furthermore, these random variables are independent in the sense that the value assumed by one has no effect on the value assumed by any of the others. The random variables $X_1, X_2, X_3, \ldots, X_n$ can be thought of as a "random sample."

Once we have actually selected n batteries for study and have observed the life span of each battery, we will have available n numbers, $x_1, x_2, x_3, \ldots, x_n$. These numbers are the observed values of the random variables $X_1, X_2, X_3, \ldots, X_n$ and can be thought of as a "random sample."

As you can see, the term "random sample" is used in three different but closely related ways in applied statistics. It may refer to the *objects* selected for study, to the *random variables* associated with the objects to be selected, or to the *numerical values* assumed by those variables. It is usually clear from the context of the discussion which is intended. These ideas are illustrated in Fig. 6.1.

Even though the term "random sample" is used in these three ways, the formal definition of the term is mathematical in nature. When we use the term in stating theoretical results we mean the following.

Definition 6.1.1 (Random sample). A random sample of size n from the distribution of X is a collection of n independent random variables each with the same distribution as X.

The theorems and definitions presented later use the term "random sample" in the sense just described. When objects are selected from a finite population, this type of sample results only when sampling is done with replacement. That is, an object is drawn, observed, and placed back in the population for possible reselection. This ensures that $X_1, X_2, X_3, \ldots, X_n$ are indeed independent and identically distributed. Usually, sampling from a finite population is done without replacement. This means that the random variables $X_1, X_2, X_3, \ldots, X_n$ are not independent. However, if the sample is small relative to the population itself, then removal of a few items does not drastically alter the composition of the population. A generally accepted guideline is that for all practical purposes we may assume independence whenever the sample constitutes at most 5% of the population. If this is not true, then the techniques used to estimate parameters must be altered to take this into account. We shall be assuming that for all practical purposes $X_1, X_2, X_3, \dots, X_n$ are independent in the discussions that follow.

Once a random sample has been drawn, we usually use the data gathered to evaluate pertinent *statistics*. What is a statistic? Roughly speaking, a statistic is a random variable whose numerical value can be determined from a random sample. That is, a statistic is a random variable that is a function of the elements of a random sample $X_1, X_2, X_3, \dots, X_n$. Typical statistics of interest to statisticians are $\sum_{i=1}^{n} X_i$, $\sum_{i=1}^{n} X_i^2$, $\sum_{i=1}^{n} X_i/n$, $\max_i\{X_i\}$, and $\min_i\{X_i\}$. These ideas are illustrated in Example 6.1.1.

Example 6.1.1. Consider the random variable X, the number of times per hour that a television signal is interrupted by random interference. Assume that this random variable has a Poisson distribution with unknown mean μ and unknown variance σ^2. To approximate the value of each of these parameters, we intend to observe the signal for ten randomly selected nonoverlapping one-hour periods over a week's time. Let X_i $(i = 1, 2, 3, \dots, 10)$ denote the number of interruptions that occur during the ith observation period. The random variables $X_1, X_2, X_3, \dots, X_{10}$ constitute a random sample size 10 from a Poisson distribution with unknown mean μ and unknown variance σ^2. When the experiment is conducted, these data result:

$$x_1 = 1 \quad x_3 = 0 \quad x_5 = 1 \quad x_7 = 0 \quad x_9 = 3$$
$$x_2 = 0 \quad x_4 = 2 \quad x_6 = 1 \quad x_8 = 0 \quad x_{10} = 0$$

The observed values of the statistics $\sum X_i$, $\sum X_i^2$, $\sum X_i/n$, $\max_i\{X_i\}$, and $\min_i\{X_i\}$ based on this sample are 8, 16, .8, 3, and 0, respectively. Note that the random variable $X_1 - \mu$ is *not* a statistic. Since μ is unknown, we cannot determine its numerical value from a random sample.

6.2 PICTURING THE DISTRIBUTION

When studying a random variable X, one important question to be answered is: "To which family of random variables does X belong?" That is, we need to determine whether X is binomial, Poisson, normal, exponential, or belongs to some other family of variables. In the discrete case it is often possible to determine the appropriate family from the physical description of the experiment. The only job left for the statistician is to approximate the values of the parameters that characterize the distribution. Continuous random variables are more difficult to handle. To determine the family to which such a variable belongs, we must get an idea of the *shape* of its density. For example, if the density appears to be flat then it is reasonable to suspect that X is uniformly distributed; if it is bell-shaped then X may be normally distributed.

Stem-and-Leaf Charts

Here we consider some graphical methods for studying the distribution of a continuous random variable. The first method entails constructing what is called a "stem-and-leaf" diagram. This method was first introduced by John Tukey in 1977 [48].

A stem-and-leaf diagram consists of a series of horizontal rows of numbers. Each row is labeled via a number called its "stem"; the other numbers in the rows are called "leaves." There are no rigid rules as to how to construct such a diagram. Basically these steps are followed:

1. Choose some convenient numbers to serve as stems. The stems are usually the first one or two digits of the numbers in the data set.
2. Label the rows via the stems selected.
3. Reproduce the data set graphically by recording the digit following the stem as a leaf.
4. Turn the graph on its side to get an idea of the shape of the distribution.

These ideas are illustrated in Example 6.2.1.

Example 6.2.1. To study the random variable X, the life span in hours of the lithium battery in a particular model of pocket calculator, a random sample of 50 batteries is obtained and the life span of each is determined. These data result:

4285	564	1278	205	3920
2066	604	209	602	1379
2584	14	349	3770	99
1009	4152	478	726	510
318	737	3032	3894	582
1429	852	1461	2662	308
981	1560	701	497	3367
1402	1786	1406	35	99
1137	520	261	2778	373
414	396	83	1379	454

To construct a "stem-and-leaf" diagram for these data, we first choose numbers to serve as "stems." It is often convenient to use the first digit of a number as its stem. If a three-digit number such as 318 is expressed as a four-digit number (0318) by including a leading zero, then this data set entails the use of the five stems $0, 1, 2, 3, 4$. We shall use the second digit of a number as its "leaf." The diagram is constructed by listing the stems as a vertical column as shown in Fig. 6.2(a). The first observation, 4285, has a stem of 4 and a leaf of 2. It is represented in the diagram as shown in Fig. 6.2(b). The entire data set, recorded in the order in which the observations appear, is shown in Fig. 6.2(c).

Is it reasonable to assume that X is normally distributed? To answer this question, turn the stem-and-leaf diagram on its side and look for the bell-shape

0	0 |		0 |	394 560 785 323 472 026 740 055 303 4
1	1 |		1 |	044 157 244 33
2	2 |		2 |	056 7
3	3 |		3 |	078 93
4	4 |	2	4 |	21
(a)	(b)			(c)

FIGURE 6.2
(a) The integers $0, 1, 2, 3, 4$ form the stems for a stem-and-leaf diagram. (b) The number 4285 has a stem of 4 and a leaf of 2. (c) Complete stem-and-leaf diagram for the sample of battery life spans of Example 6.2.1.

characteristic of a normal density. This bell shape is not present, leading us to suspect that X is *not* a member of the family of normal random variables.

Histograms and Ogives

The stem-and-leaf diagram provides a quick but rough look at the data set. A second and more thorough method for categorizing data is now considered. It entails breaking the data into "classes" or categories, determining the number of observations in each class, and constructing a graph to display these frequencies. This graph, called a *histogram*, is a vertical bar graph. It depicts the frequency distribution using bars constructed so that the area of each bar is proportional to the number of objects in the respective category.

We illustrate this method using the data of Example 6.2.1. Our task is to separate these data into categories. Usually, 5 to 20 categories of equal length are desirable, with the actual number used being dependent on the number of data points available. Since we have only 50 observations, we use a relatively small number of categories, say seven. Now we locate the largest data point (4285) and the smallest (14). These are used to find the length of the interval containing all of the data points. In this case, the data are covered by an interval of length $4285 - 14 = 4271$ units. To find the minimum length required for each category, this number is divided by the number of categories desired. Here the minimum category length is $4271/7 \doteq 610.14$ units. To find the actual category length to be used in splitting the data, we round *up* the minimum length to the same number of decimal places as the data. Here the data are reported in whole numbers. Thus we round up the minimum length, 610.14, to the nearest whole number, 611. The categories actually used will be of length 611. The first category starts 1/2 unit below the smallest observation. Since the data here are integer-valued, a unit is one and we start the first category 1/2 unit $= 1/2 \times 1 = .5$ below the smallest observation. That is, the lower boundary for the first category is $14 - .5 = 13.5$. The remaining category boundaries are found by successively adding the actual category length (611) to the preceding boundary until all data points are covered. In this manner we obtain the following seven finite categories for the battery lives: 13.5 to 624.5, 624.5 to 1235.5, 1235.5 to 1846.5, 1846.5 to 2457.5, 2457.5 to 3068.5, 3068.5 to 3679.5, and 3679.5 to 4290.5. Note that since the boundaries have one more decimal place than the data, no data point can fall on a boundary; each data point must fall into exactly one category. The data can be summarized now in table form by recording the number (frequency) and the percentage (relative frequency) of the observations in each category as shown in Table 6.1. From this table we can construct a histogram of the data. If the frequency per category is plotted along the vertical axis, the resulting bar graph is called a *frequency histogram*; if the vertical axis is used to plot the relative frequency per category, then the diagram is called a *relative frequency histogram*. Both plots provide a visual display of the data that conveys an idea of the shape of the density of the random variable X under study. The relative frequency histogram for the data of Example 6.2.1 is shown in Fig. 6.3. Since the histogram does not

TABLE 6.1

Category	Boundaries	Frequency	Relative frequency
1	13.5 to 624.5	23	23/50 = 46%
2	624.5 to 1235.5	7	7/50 = 14%
3	1235.5 to 1846.5	9	9/50 = 18%
4	1846.5 to 2457.5	1	1/50 = 2%
5	2457.5 to 3068.5	4	4/50 = 8%
6	3068.5 to 3679.5	1	1/50 = 2%
7	3679.5 to 4290.5	5	5/50 = 10%

exhibit a bell shape, we see once again that these data do not support an assumption of normality.

In addition to the frequency distribution among categories, it is of interest to consider the cumulative frequency distribution of the observations. The cumulative frequency distribution is found by determining for each category the number and percentage of observations falling in or below that category. The cumulative distribution of the data of Example 6.2.1 is shown in Table 6.2.

When the random variable under study is continuous, the cumulative distribution can be used to construct a graph that approximates its cumulative distribution function F. The graph is a line graph obtained by plotting the upper boundary of each category on the horizontal axis against the relative cumulative

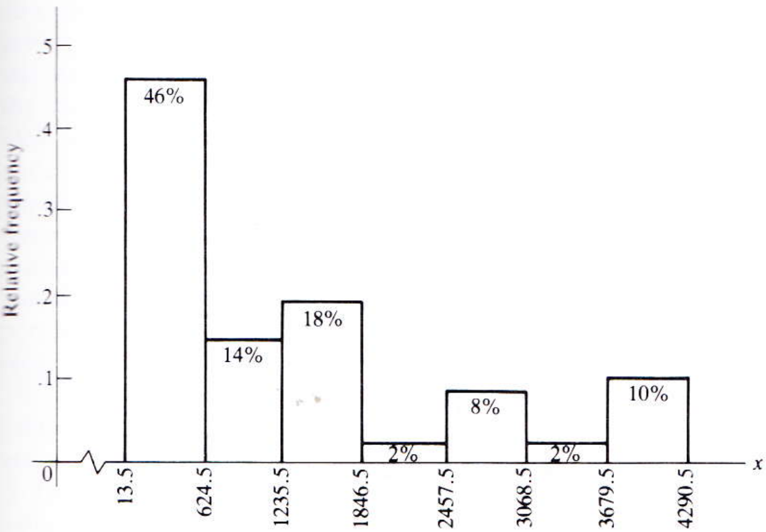


FIGURE 6.3
Relative frequency histogram for the sample of battery life spans of Example 6.2.1.

TABLE 6.2

Category	Boundaries	Frequency	Cumulative frequency	Relative cumulative frequency
1	13.5 to 624.5	23	23	$23/50 = 46\%$
2	624.5 to 1235.5	7	30	$30/50 = 60\%$
3	1235.5 to 1846.5	9	39	$39/50 = 78\%$
4	1846.5 to 2457.5	1	40	$40/50 = 80\%$
5	2457.5 to 3068.5	4	44	$44/50 = 88\%$
6	3068.5 to 3679.5	1	45	$45/50 = 90\%$
7	3679.5 to 4290.5	5	50	$50/50 = 100\%$

frequency. This type of graph is called a *relative cumulative frequency ogive*. The ogive for the data of Example 6.2.1 is shown in Fig. 6.4. From the ogive we can answer questions such as these: "Approximately what percentage of batteries fail during the first 1500 hours of operation?" "What time represents the midway point in the sense that half the batteries fail on or before this time?"

The first question can be answered graphically by locating 1500 on the horizontal axis, projecting a vertical line up to the ogive, and then projecting a horizontal line over to the vertical axis as shown in Fig. 6.5. The desired percentage is seen to be approximately 70%. The second question is answered by locating .5 on the vertical axis and reversing the process. The answer is seen to be approximately 850 hours. (See Fig. 6.5.)

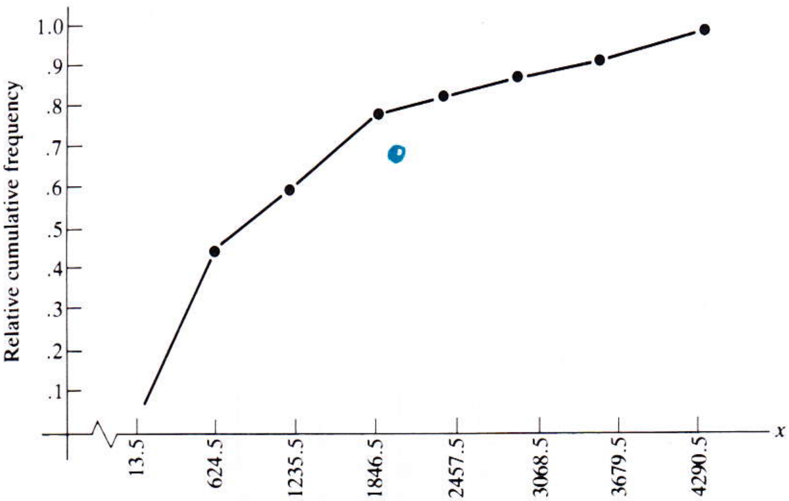


FIGURE 6.4
Relative cumulative frequency ogive for the sample of battery lifespans of Example 6.2.1.

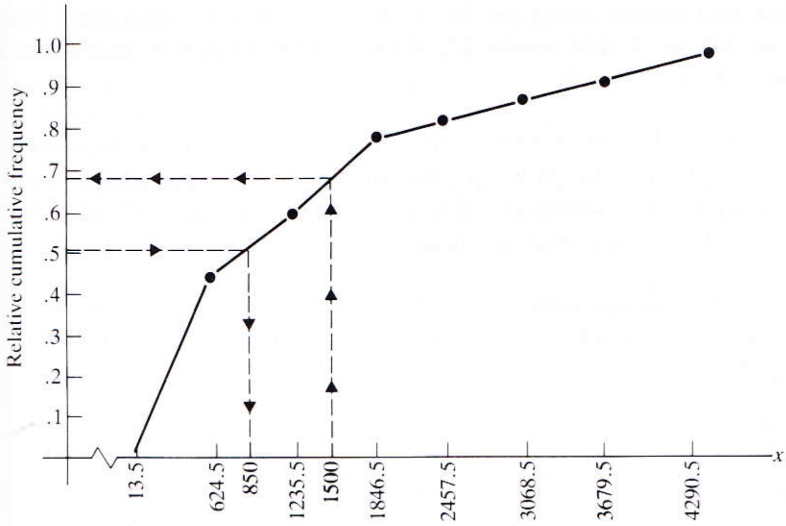


FIGURE 6.5
Projective method of approximating probabilities using a relative cumulative frequency ogive.

6.3 SAMPLE STATISTICS

We have seen that the behavior of a random variable, X, is determined by its density. We have also seen that the parameters μ, the theoretical average value of the random variable, and σ^2, its variability about the mean, are helpful in describing X. In the last section, we considered some graphical methods for getting an idea of the shape of the density. In this section, we consider some statistics that allow us to summarize a data set analytically. Since it is hoped that the data set reflects the population as a whole, these statistics also give us some idea of the values of the parameters that characterize X over the population under study. In particular, we consider two measures of location or central tendency in a data set, the *sample mean* and the *sample median*. We also consider three measures of variability within the data set, the *sample variance*, *sample standard deviation*, and *sample range*. The word sample is used to emphasize the fact that the data sets presented are based on experiments involving only a small portion of objects that constitute the population being studied. That is, they represent a random sample from the distribution of X.

The mean or theoretical average value of X is our primary measure of the center of location of X. The primary measure of the center of location of a data set is its arithmetic average. Since we view a data set as a set of observations on X, the arithmetic average for a particular set of observations is just the observed value of the *statistic* $\sum_{i=1}^{n} X_i/n$. This statistic, called the *sample mean* is defined formally in the next definition.

Definition 6.3.1 (Sample mean). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from the distribution of X. The statistic $\sum_{i=1}^{n} X_i/n$ is called the sample mean and is denoted by $\overline{X}$.

Note that μ_X and $\overline{X}$ are *not* the same. The parameter μ_X is the theoretical average value for X over the entire population; $\overline{X}$ is a statistic which, when evaluated over a particular random sample, gives the average value of X *for that sample*. It is hoped, of course, that the observed value of $\overline{X}$ is close to μ_X.

Example 6.3.1. A random sample of size 9 yields the following observations on the random variable X, the coal consumption in millions of tons by electric utilities for a given year:

406	395	400	450	390	410	415	401	408

The observed value of the sample mean for these data is

$$\bar{x} = \sum_{i=1}^{n} x_i/n = (406 + 395 + 400 + \cdots + 408)/9$$

$$= 3675/9 \doteq 408.3 \text{ million tons}$$

The average value for X for this sample is 408.3 million tons. What is the average number of tons of coal used by electric utilities across the country in this particular year? That is, What is μ_X? Unfortunately, this question cannot be answered with certainty from this sample. However, the sample leads us to believe that μ_X lies close to 408.3 million tons. Admittedly, the word "close" is a bit vague. In Chap. 8 we shall consider a method for determining how close μ_X is likely to be to 408.3 million tons.

A second measure of the center of location of a random variable X is its *median*. The median of a random variable is its 50th percentile (see Exercise 12). That is, the median for X is that number M such that

$$P[X < M] \leq .50 \qquad \text{and} \qquad P[X \leq M] \geq .50$$

If X is continuous, then its median is the "halfway point" in the sense that an observation on X is just as likely to fall below M as it is to fall above it. We define the median for a sample with this in mind.

Definition 6.3.2 (Sample median). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n arranged in order (smallest to largest). The sample median, denoted by $\tilde{X}$, is given by

$$\tilde{X} = \begin{cases} X_{(n+1)/2} & \text{if } n \text{ is odd} \\ \dfrac{(X_{(n/2)} + X_{(n/2)+1})}{2} & \text{if } n \text{ is even} \end{cases}$$

This definition appears to be complicated. It is not. It says simply that to find the sample median, first the observations are ordered smallest to largest. The sample median is the middle observation if n is odd; it is the arithmetic average of the two middle observations if n is even.

Example 6.3.2. The nine observations on X, the coal consumption in millions of tons by electric utilities for a given year, arranged in order are

390	395	400	401	406	408	410	415	450

Since $n = 9$ is odd, the median for this sample is $x_{(n+1)/2} = x_5$. This observation, 406, is the middle value in our ordered list. Note that this is the median for this data set. It gives us a *rough* idea of the median coal consumption across the country during the year.

Recall that we are usually concerned not only with the mean of a random variable but also with its variance. The variance of a random variable, given by

$$\sigma^2 = E\left[(X - \mu)^2\right]$$

measures the variability of X about the population mean. We want to develop an analogous measure of variability within a sample. To do so, we parallel the logic used in defining σ^2. We do not know the value of the population mean, but we will have available an observed value for the sample mean. We cannot observe the differences $(X - \mu)^2$ for all members of the population, but we can observe the difference $(X_i - \overline{X})^2$ for each element X_i of the random sample. Since σ^2 is an expectation, a theoretical average value, logic dictates that we replace this operation by an arithmetic average of sample values. That is, the natural measure of variability within a sample that parallels our definition of variability within the population is

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n}$$

This method of measuring variability within a sample is acceptable. In fact, many electronic calculators with built-in statistical capability utilize this formula to compute the variance of a sample. However, in most cases we shall be using the variability in the sample to approximate σ^2. This measure of variability lacks one important property that is usually considered desirable for such purposes. In particular, it is not *unbiased* for σ^2. It can be shown that if we divide $\sum_{i=1}^{n}(X_i - \overline{X})^2$ by $n - 1$ rather than n, then the resulting measure will be unbiased. (See Chap. 7.) For this reason, we choose to define the variance of a sample as given in Definition 6.3.3. The definition of the term *sample standard deviation* follows logically.

Definition 6.3.3 (Sample variance and sample standard deviation). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from the distribution of X. Then the statistic

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

is called the *sample variance*. Furthermore, the statistic $S = \sqrt{S^2}$ is called the *sample standard deviation*.

Recall that when we computed the value of σ^2 in Chap. 3, the actual definition of the term "variance" was seldom used; a computational formula was developed that was arithmetically easier to handle than the definition. The same is true here. When S^2 is evaluated from a sample, Definition 6.3.3 is not usually used. Rather, we use a computational formula.

Theorem 6.3.1 (A computational formula for S^2). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from the distribution of X. The sample variance is given by

$$S^2 = \frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i \right)^2}{n(n-1)}$$

The proof of this theorem is a direct consequence of the rules of summation and is left as an exercise. (Exercise 23.) To illustrate the use of the theorem, we calculate the sample variance and sample standard deviation for the data of Example 6.3.2.

Example 6.3.3. These data constitute a sample of observations on X, the coal consumption in millions of tons by electric utilities for a given year:

390	400	406	410	450	395	401	408	415

To compute the sample variance we must evaluate the statistics $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{n} X_i^2$ for this sample. The observed values are

$$\sum_{i=1}^{9} x_i = 3675 \qquad \sum_{i=1}^{9} x_i^2 = 1{,}503{,}051$$

The observed value of S^2 is

$$s^2 = \frac{9 \sum_{i=1}^{9} x_i^2 - \left(\sum_{i=1}^{9} x_i \right)^2}{9(8)} = \frac{9(1{,}503{,}051) - (3675)^2}{9(8)} \doteq 303.25$$

The observed value of S is

$$s = \sqrt{s^2} = \sqrt{303.25} \doteq 17.4 \text{ million tons}$$

Note that 17.4 million tons is the standard deviation for this sample. It is not the standard deviation in coal consumption for all electric utilities across the country for the given year. However, it does indicate that σ probably has a value close to 17.4 million tons.

The last sample statistic to be considered is the *sample range*. This statistic was used in categorizing data in Sec. 6.2 even though the word "range" was not mentioned at the time. The sample range is defined to be the difference between the largest and smallest observations. Thus the sample range for the data of Example 6.3.3 is $450 - 390 = 60$ million tons.

One word of caution is in order. We have assumed that the data set presented in this section represents a random sample drawn from a larger population because this is the situation most often encountered in practice. Occasionally you will encounter a data set that is *not* a sample. Rather, it represents an observation on X for *every* member of the population. If this is the case, then the population mean is just the arithmetic average of these observations; that is, $\mu = \bar{x}$. Furthermore, the population variance is given by

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

Be careful. Be sure that you understand the nature of your data set before you begin to summarize its properties.

CHAPTER SUMMARY

This chapter is a link between the study of probability in its own right and the use of probability in the study of applied statistics. We began by defining exactly what we meant by the term "random sample." In particular we noted that the term is used in three ways. It can denote the objects sampled, the random variables associated with those objects, or the numerical values assumed by these random variables. We noted also that in this text we are assuming either that sampling is from an infinite population, sampling is done with replacement from a finite population, or sampling without replacement from a finite population is done in such a way that the sample constitutes at most 5% of the population. This ensures that it is reasonable to assume that the random variables $X_1, X_2, \ldots, X_n$ are, for all practical purposes, independent. We introduced two graphical methods for picturing the distribution of a data set. These methods, the stem-and-leaf chart and histograms, help determine the type of random variable with which we are dealing. That is, they help us get an idea of the shape of the density f associated with the random variable. The relative cumulative frequency ogive was introduced as a means of approximating the cumulative distribution function, F, of a continuous random variable. We introduced some summary statistics that serve two purposes. They describe the data set at hand and they help approximate the value of corresponding parameters associated with the population from which the sample was drawn. These terms were introduced.

Population	Sample mean
Percentile	Random sample
Median	Quartile
Statistic	Sample median
Decile	Stem and leaf
Sample variance	Interquartile range
Frequency histogram	Sample standard deviation
Relative frequency histogram	Sample range
Relative cumulative frequency ogive	Outlier

EXERCISES

Section 6.1

In Exercises 1 through 5, a problem is described. In each case, decide whether a statistical study is appropriate. If so, explain why you think this is the case and identify the population(s) of interest.

1. A bridge is to be built across a deep canyon. An engineer is interested in determining the distribution of the random variable X, the maximum wind speed per day at the site so that the bridge can be designed to withstand potential stresses that will be placed upon it from this source.
2. A botanist thinks that indoleacetic acid is effective in stimulating the formation of roots in cuttings from lemon trees. In an experiment to verify this contention, two groups of cuttings are to be used. One group is to be treated with a dilute solution of indoleacetic acid; the other is given only water. Later a comparison of the root systems of the two groups will be made.
3. An architectural firm is to sublet a contract for a wiring project. Seven electrical contractors are available for the job. We want to determine the average cost of the job and the average time required to complete the job for these seven contractors.
4. A computer system has a number of remote terminals attached to it. To decide whether or not to increase this number, it is necessary to study the random variable X, the length of time expended per session by users of the terminals currently in place.
5. Prior to changing from the traditional eight-hour-a-day, five-day-a-week work schedule to a ten-hour-a-day, four-day-a-week schedule, the opinion of the 50,000 workers who would be affected is to be sought.
6. Air quality is of concern to everyone. It is judged by the number of micrograms of particulate present per cubic meter of air. Assume that this variable is normally distributed with unknown mean and unknown variance. Monitoring stations sample air by sucking it through a thin fiberglass sheet that collects the fine particles suspended in the air. In a particular locality this is done for five randomly selected 24-hour periods each month. Thus, each month a random sample of size $n = 5$ from

a normal distribution is available.

(*a*) Consider the random variable X_1, the particulate level for the first 24-hour period studied during a given month. What is the distribution of this random variable?

(*b*) For given month, these readings result:

$$x_1 = 45 \qquad x_2 = 50 \qquad x_3 = 62 \qquad x_4 = 57 \qquad x_5 = 70$$

For these data, evaluate the statistics ΣX_i, ΣX_i^2, $\Sigma X_i/n$, $\max_i\{X_i\}$, $\min_i\{X_i\}$.

(*c*) Is the random variable $X_5 - \mu$ a statistic? Is the random variable $(X_5 - \mu)/\sigma$ a statistic? Explain.

Section 6.2

7. A data set containing 70 observations each reported to one decimal place is to be split into eight categories. The largest observation is 75.1 and the smallest is 16.3.

(*a*) These data are covered by an interval of what length?

(*b*) Using the method outlined in this section, each category will be of what length?

(*c*) Since these data are reported to the nearest 1/10, a unit is 1/10 and a half unit is $1/2 \cdot 1/10 = 1/20 = .05$. What is the lower boundary for the first category?

(*d*) What are the boundaries for each of the eight categories?

8. Acute exposure to cadmium produces respiratory distress, kidney and liver damage, and may result in death. For this reason, the level of airborne cadmium dust and cadmium oxide fume in the air is monitored. This level is measured in milligrams cadmium per cubic meter of air. A sample of 35 readings yields these data: (Based on a report in *Environmental Management*, September 1981, p. 414.)

.044	.030	.052	.044	.046
.020	.066	.052	.049	.030
.040	.045	.039	.039	.039
.057	.050	.056	.061	.042
.055	.037	.062	.062	.070
.061	.061	.058	.053	.060
.047	.051	.054	.042	.051

(*a*) Construct a stem-and-leaf diagram for these data. Use the numbers 02, 03, 04, 05, 06, and 07 as stems.

(*b*) Would you be surprised to hear someone claim that the random variable X, the cadmium level in the air, is normally distributed? Explain.

(*c*) Use the method outlined in this section to break these data into six categories. (Here a unit is 1/1000 and a half unit is .0005.)

(*d*) Construct a frequency table and a relative frequency histogram for these data. Does the histogram exhibit the bell shape characteristic of a normal density?

(*e*) Construct a cumulative frequency table and a relative cumulative frequency ogive for these data. Use the ogive to approximate that point above which 50% of the readings should fall.

9. Consider the stem-and-leaf diagram of Fig. 6.2. What family of distributions is suggested by the diagram?

10. Liquid products were first obtained from coal in England during the 1700s. Lamp oil was produced from coal in the United States as early as 1850 but the domestic coal chemicals industry did not develop until World War I. A modern coal-for-recovery system uses a battery of coke ovens to produce liquid products from the coal feed. These observations are obtained on the random variable X, the number of gallons of liquid product obtained per ton of coal feed. (Based on a report in *McGraw-Hill Yearbook of Science and Technology*, 1983, p. 37.)

7.6	8.2	7.1	10.0	6.5	9.6
6.1	6.2	7.6	6.2	9.5	6.7
7.4	9.5	9.2	8.0	8.5	9.3
8.8	9.6	9.7	6.8	7.1	7.7
8.7	7.8	8.7	8.2	8.2	7.4
9.0	8.8	7.3	7.9	7.1	7.9
7.6	6.7	8.1	6.2	5.3	7.4
7.7	9.1	7.9	8.7	8.4	8.1

(*a*) Construct a stem-and-leaf diagram for these data. Use the numbers 5, 6, 7, 8, 9, 10 as stems.
(*b*) Is the assumption that X is normally distributed justifiable? Explain.
(*c*) Use the method outlined in this section to break these data into seven categories.
(*d*) Construct a frequency table and a relative frequency histogram for these data. Does the histogram exhibit the bell shape characteristic of a normal density?
(*e*) Construct a cumulative frequency table and a relative cumulative frequency ogive for these data. Use the ogive to approximate the probability that a randomly selected ton of coal will yield less than seven gallons of liquid product.

11. Some efforts are currently being made to make textile fibers out of peat fibers. This would provide a source of cheap feedstock for the textile and paper industries. One variable being studied is X, the percentage ash content of a particular variety of peat moss. Assume that a random sample of 50 mosses yields these observations: (Based on data found in "Peat Fibre: Is There Scope in Textiles?" *Textile Horizons*, vol. 2, no. 10, October 1982, p. 24.)

.5	1.8	4.0	1.0	2.0
1.1	1.6	2.3	3.5	2.2
2.0	3.8	3.0	2.3	1.8
3.6	2.4	.8	3.4	1.4
1.9	2.3	1.2	1.9	2.3
2.6	3.1	2.5	1.7	5.0
1.3	3.0	2.7	1.2	1.5
3.2	2.4	2.5	1.9	3.1
2.4	2.8	2.7	4.5	2.1
1.5	.7	3.7	1.8	1.7

(*a*) Construct a stem-and-leaf diagram for these data. Use the numbers 0, 1, 2, 3, 4, 5 as stems.
(*b*) Is there any reason to suspect that X is not normally distributed? Explain.
(*c*) Use the method outlined in this section to break these data into seven categories.

(*d*) Construct a frequency table and a relative frequency histogram for these data. Does the histogram suggest that X might not be normally distributed? If so, what distribution might be appropriate?

(*e*) Construct a cumulative frequency table and a relative cumulative frequency ogive for these data. Use the ogive to approximate the probability that a randomly selected specimen of this variety of moss will have an ash content that exceeds 2%.

12. (*Percentiles.*) Let X be a random variable. The point $p_{k/100}$ $(k = 1, 2, 3, \ldots, 100)$ such that

$$P\left[X < p_{k/100}\right] \leq k/100 \qquad \text{and} \qquad P\left[X \leq p_{k/100}\right] \geq k/100$$

is called the *kth percentile* for X. For example, let X be binomial with $n = 20$ and $p = .5$. The 25th percentile for X is the point $p_{25/100} = 8$ since, from Table I of App. A, we see that

$$P[X < 8] = .1316 \leq .25 \qquad \text{and} \qquad P[X \leq 8] = .2517 \geq .25$$

(*a*) Let X be binomial with $n = 20$ and $p = .5$. Find the 60th percentile for X.

(*b*) Let X be Poisson with $\lambda s = 10$. Find the 30th percentile for X.

(*c*) Argue that in the case of a continuous random variable the kth percentile is that point such that $P[X \leq k/100] = k/100$.

(*d*) Let X be exponentially distributed with $\beta = 1$. Show that the 20th percentile for X is $-\ln .80$. *Hint:* Find the point p such that

$$\int_0^p e^{-x}\, dx = .20.$$

13. (*Quartiles.*) The 25th, 50th, 75th, and 100th percentiles for X are called its first, second, third, and fourth *quartiles* respectively.

(*a*) State the definition of the first quartile in terms of probabilities.

(*b*) Let X be binomial with $n = 20$ and $p = .5$. Find the first quartile for X.

(*c*) Let X be exponentially distributed with $\beta = 1$. Find the first quartile for X.

14. (*Deciles.*) The 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, and 100th percentiles for X are called its *deciles*.

(*a*) State the definition of the 40th decile for X in terms of probabilities.

(*b*) Let X be Poisson with $\lambda s = 10$. Find the 6th decile for X.

(*c*) Let X be exponentially distributed with $\beta = 1$. Find the third decile for X.

15. The percentiles, quartiles, and deciles for a continuous random variable can be approximated from a relative cumulative frequency ogive using the projective method. For instance, in Fig. 6.5 we approximated the 50th percentile for X, the life span of a lithium battery, to be 850 hours.

(*a*) Approximate the first quartile for X, the cadmium level in the air, using the data of Exercise 8.

(*b*) Approximate the fourth decile for X, the number of gallons of liquid product obtained per ton of coal fuel, using the data of Exercise 10.

(*c*) Approximate the 50th percentile for X, the percentage ash content for a particular variety of moss using the data of Exercise 11.

16. When running computer programs on a time-sharing basis, costs vary from session to session. These observations are obtained on the random variable X, the cost per session to the user.

\$1.08	.84	1.41	.99	.82
.89	.38	1.05	1.19	.65
1.09	1.03	.81	.55	.71
1.89	.47	.59	1.22	1.27
1.02	1.09	1.02	.86	1.23
1.23	.85	1.02	1.25	.80

Construct a relative cumulative frequency ogive for these data. Use the ogive to approximate the 50th percentile; the first quartile; the third quartile. The *interquartile range* is defined by to be the difference between the third and first quartiles. Approximate the interquartile range for X based on these data.

Section 6.3

17. Consider these data sets:

I			II			
1	3	2	1	2	4	1
2	5	4	2	5	2	5
4	3	3	1	5	5	3

(*a*) Find the sample mean and sample median for each data set.
(*b*) Find the sample range for each data set.
(*c*) Find the sample variance and sample standard deviation for each data set.
(*d*) Would you be surprised to hear someone claim that these data were drawn from the same population? Explain. *Hint:* Consider the shape of the distribution as well as the observed values of the sample statistics.

18. The observed values of the statistics $\Sigma_{i=1}^{50} X_i$ and $\Sigma_{i=1}^{50} X_i^2$ for the data of Example 6.2.1 are $\Sigma_{i=1}^{50} x_i = 63{,}707$ and $\Sigma_{i=1}^{50} x_i^2 = 154{,}924{,}261$.
(*a*) Would you be surprised to hear someone claim that the mean life span of the lithium batteries used in this model calculator is 1270 hours? Explain.
(*b*) Find the sample variance and sample standard deviation for these data.

19. Use the data of Example 6.1.1 to approximate the mean and variance of the random variable X, the number of times per hour that a television signal is interrupted by random interference.

20. Use the data of Exercise 8 to approximate the mean, variance, and standard deviation of the random variable X, the level of airborne cadmium dust and cadmium oxide fume. Assume that these approximations are fairly accurate. Between what two values would you expect approximately 95% of the readings to fall? Explain.

21. Use the data of Exercise 10 to approximate the mean, variance, and standard deviation of the random variable X, the number of gallons of liquid product obtained per ton of coal feed.

22. Use the data of Exercise 11 to approximate the mean, variance, and standard deviation of the random variable X, the percentage ash content of a particular variety of peat moss.

23. Prove Theorem 6.3.1. *Hint:* Begin by expanding the term $(X_i - \bar{X})^2$. Apply the rules of summation and remember that relative to summation on i, $\bar{X}$ is constant.

24. (*Outliers.*) Temperature differences between the warm upper surface of the ocean and the colder deeper levels can be utilized to convert thermal energy to mechanical energy. This mechanical energy can in turn be used to produce electrical power using a vapor turbine. Let X denote the difference in temperature between the surface of the water and the water at a depth of 1 kilometer. Measurements are taken at 15 randomly selected sites in the Gulf of Mexico. These data result: (Based on a report in *McGraw-Hill Yearbook of Science and Technology*, 1983, p. 410.)

22.5	23.8	23.2	22.8	10.1*
23.5	24.0	23.2	24.2	24.3
23.3	23.4	23.0	23.5	22.8

(*a*) Find the sample mean, sample median, and sample standard deviation for these data.

(*b*) Note that the starred observation in the data set is very different from the others. It is called an *outlier*. (An outlier is an observation at either extreme of a sample which is so far removed from the main body of the data that the appropriateness of including it in the sample is questionable.) Some apparent outliers are actually misrecorded data points; these can be found and corrected. Other outliers are legitimate observations that just happen to be far removed from the body of the data; these should not be ignored. To see the effect of this outlier, drop it from the data set and calculate the sample mean, median, and standard deviation for the remaining 14 observations. Which measure is least affected by the presence of the outlier? Do you see why it is desirable to report both the mean and median of a data set?

*25. (*Approximating σ via the range.*) The range can play an important role in the design of statistical studies. To obtain a prespecified degree of accuracy when estimating population parameters, an adequate sized sample must be drawn. Most formulas used to determine sample size require knowledge of σ, the population standard deviation. Often the researcher will not have an estimate of σ available but will have an idea of the expected range of his or her data. In Sec. 4.5 we saw that when sampling from a normal distribution

$$P[-2\sigma < X - \mu < 2\sigma] \doteq .95$$

If X is not normally distributed, then Chebyshev's inequality can be applied to conclude that

$$P[-3\sigma < X - \mu < 3\sigma] \geq .89$$

That is, X always lies within at most 3 standard deviations of its mean with high probability. From this, it can be concluded that the estimated range covers an interval of roughly 4σ for normally distributed random variables and 6σ otherwise. In the normal case, an estimate of σ can be obtained by solving the equation

$$4\sigma \doteq \text{estimated range}$$

for σ. Thus we see that

$$\sigma \doteq (\text{estimated range})/4$$

when X is normally distributed. If X is not normally distributed, then

$$\sigma \doteq (\text{estimated range})/6$$

These data are obtained on the random variable X, the cpu time in seconds required to run a program using a statistical package.

6.2	5.8	4.6	4.9	7.1	5.2
8.1	.2	3.4	4.5	8.0	7.9
6.1	5.6	5.5	3.1	6.8	4.6
3.8	2.6	4.5	4.6	7.7	3.8
4.1	6.1	4.1	4.4	5.2	1.5

(*a*) Construct a stem-and-leaf diagram for these data. Is the assumption justified that X is normally distributed?
(*b*) Approximate σ via the sample standard deviation s.
(*c*) Find the sample range for these data and use it to approximate σ. Compare your result to that obtained in part (*b*).

REVIEW EXERCISES

26. Bricks are produced in lots of size 1000. Before shipping a lot, a sample of 25 bricks is selected and inspected for quality. Two random variables are of interest. These are X, the number of chips per brick, and Y, the hardness of the brick. Assume that hardness is measured on a continuous scale from 1 to 10 with larger numbers indicating a harder brick.

		x					y		
2	5	0	1	2	3.2	6.3	6.4	6.7	7.3
0	3	0	0	2	7.1	5.4	4.6	5.8	9.1
1	1	0	1	3	7.7	6.1	8.1	5.9	6.2
2	1	1	7	4	6.0	6.8	7.2	6.3	8.2
0	2	3	5	1	5.1	4.2	6.9	4.5	5.0

(*a*) What is the name of the family of random variables to which X belongs?
(*b*) Approximate the mean, variance, standard deviation, and median of X based on these data.
(*c*) Construct a stem-and-leaf diagram for the hardness measurements. Based on this diagram, would it be unrealistic to assume that Y is approximately normally distributed?
(*d*) Approximate the mean, variance, standard deviation, and median of Y.

27. In an attempt to study the problem of failure in field installed computer equipment, data is collected on fifty field trips made to repair equipment. The random variables studied are X, the time in hours required to locate and rectify the problem, and Y, the cause of the failure. We define Y by

$$Y = \begin{cases} 1 & \text{if the failure is due to a faulty microprocessor chip} \\ 0 & \text{otherwise} \end{cases}$$

These data are obtained:

x							y						
1.52	1.83	2.25	4.73	2.89	1.49	1.34	0	0	0	0	0	0	1
2.15	2.66	2.79	1.35	1.54	4.59	4.27	0	0	0	0	0	0	0
3.91	2.76	3.03	3.52	5.97	1.45		0	0	0	0	0	0	
3.07	2.18	1.38	2.04	1.49	1.11		1	0	0	0	0	0	
1.24	4.84	2.82	3.16	4.58	3.28		0	0	0	0	0	0	
1.30	3.01	1.20	3.42	1.86	3.49		0	0	0	0	0	0	
3.93	2.56	2.63	5.60	4.60	5.34		0	0	0	0	0	0	
1.62	2.82	4.88	2.04	1.62	.24		0	0	1	0	0	0	

(*a*) Construct a relative frequency histogram for the data on the time required to locate and rectify the problem. Use seven categories. Based on this histogram, would you be surprised to hear someone claim that X is approximately normally distributed? Explain.

(*b*) Approximate the mean, variance, and standard deviation for X.

(*c*) Construct a relative cumulative frequency ogive. Use this ogive to approximate the median for X. Approximately what percentage of problems can be located and rectified in 1.5 hours or less?

(*d*) Let p denote the probability that the failure is due to a faulty microprocessor chip. Assume that even though p is unknown its value is the same for each trip. Theoretically, Y follows a point binomial distribution with parameter p. What is the theoretical mean for Y? Approximate this mean based on these data. If asked to approximate the probability that a future failure is due to the failure of a microprocessor chip, what would you say?

(*e*) What is the theoretical variance for Y? Use your answer to part (*d*) to approximate the variance of Y. Use the sample variance to approximate σ_Y^2. Did you get the same result? Which answer is unbiased for σ_Y^2?

*(*f*) Use the technique of Exercise 25 to estimate σ_X. Compare your answer to that of part (*b*).

COMPUTING SUPPLEMENT

I. Summary Statistics and Histograms

As can be seen, the calculations necessary to summarize even a relatively small data set can become tedious and time-consuming. To alleviate this problem, several computer systems for data analysis have been developed in recent years. Among the systems in widespread use are SPSS (Statistical Package for the Social Sciences, McGraw-Hill), BMD (Biomedical Computer Programs, University of California Press), MINITAB (Pennsylvania State University), and SAS (Statistical Analysis System, SAS Institute Inc.). To use any such system, one needs little background in computer science.

We present here a very brief introduction to SAS programming to give you some experience with computer packages. Once this experience has been gained, it is not difficult to adjust to any of the other packages, for they are similar. We introduce SAS by presenting some sample programs that could be modified to analyze any of the data sets presented in this chapter. You should consult the appropriate expert at your own installation to determine the job cards necessary to access SAS.

We begin by writing a program to obtain selected sample statistics and a histogram for the data of Example 6.2.1. The variable studied is X, the life span in hours of a lithium calculator battery.

The first step in analyzing data via the SAS package is to read the data and get them into an SAS data set. This is done by means of a series of statements which name the data set (the DATA statement), describe the arrangement of the data (the INPUT statement), and signal the beginning of the data lines (the CARDS statement). SAS statements may begin in any column and always end with a semicolon.

The name chosen for the data set should be a one-word name up to eight characters long. The first character must be a letter or an underscore. Later characters can be letters, characters, or underscores. It is usually helpful to choose a name that is in some way related to the data itself. To name the data set, the SAS key word DATA is used, followed by the name chosen. For instance, in our example we might choose to name the data set "battery." We would inform the computer of this choice by typing the statement

```
DATA BATTERY;
```

on a single card (if card input is used) or on a single line (if the program is entered via a terminal).

The next statement is the INPUT statement. This statement names the variables and describes the order in which they will appear on the card or data line. Variable names are chosen at the discretion of the programmer, but they must satisfy the same guidelines as those used in naming the data set. In our example, there is only one variable, the lifespan of the battery. The input statement is simple in this case. We need only write

```
INPUT LIFESPAN;
```

This tells the computer that each card or data line will contain the value of only one variable, whose name is "life span."

The INPUT statement is followed by the CARDS statement. The purpose of this statement is to signal SAS that the data follows immediately. This statement is written

```
CARDS;
```

The data follow immediately with one observation per line. Since SAS recognizes the end of the data when it sees a semicolon, make sure that the first line after the data contains a semicolon. This can be done by entering the line containing only a semicolon. Thus far our program looks like this:

```
DATA BATTERY;          names the data set
INPUT LIFESPAN;        names the variable
CARDS;                 signals beginning of data lines
     4285
     2066              data (one observation per line)
     2584
       :
     454
;                      signals the end of data
```

Now the data are in an SAS data set named "battery." We must tell the computer what to do with the data, by means of one or more "procedure" statements. The procedure statement begins with the SAS key word PROC, followed by the name of the procedure desired. Most of the sample statistics introduced in this chapter can be obtained by using the "means" procedure. The key word for this procedure is MEANS, followed by the key words for those sample statistics desired. Some of these key words are as follows:

MAXDEC = *n*	*n* is an integer from 0 to 8 specifying the number of decimal places that will be used to print the results
MEAN	mean
STD	standard deviation
MIN	smallest value
MAX	largest value
RANGE	range
VAR	variance
STDERR	standard error of the mean
CV	coefficient of variation
N	number of observations
SUM	sum of observations
USS	sum of the squares of the observations

To find the sample statistics for our example and to have the results reported to two decimal places we write

```
PROC MEANS MAXDEC = 2 MEAN VAR STD RANGE MAX MIN
                      SUM USS N;
```

A relative frequency histogram can be obtained by calling on the "chart" procedure. The statements required are

```
PROC CHART;
VBAR LIFESPAN/TYPE = PERCENT;
```

The first statement calls for a chart to be made; the second requests a vertical bar chart of the relative frequency distribution for the variable "life span."

One further statement is desirable. This is a "title" statement that enables the programmer to print a title at the top of each page of output. The titles can contain up to 132 characters. The key word is TITLEn, where n gives the line number at which the title is to be printed. For instance, the statements

```
TITLE1   ANALYSIS;
TITLE2   OF;
TITLE3   BATTERY DATA;
```

would result in the word "analysis" being printed on line 1 of each page of output, the word "of" being printed on line 2 of each page, and the words "battery data" being printed on line 3. The entire program is as follows;

```
DATA BATTERY;
INPUT LIFESPAN;
CARDS;
    4285
    2066
    2584
      :
     454
;
TITLE1   ANALYSIS;
TITLE2   OF;
TITLE3   BATTERY DATA;
PROC MEANS MAXDEC = 2 MEAN VAR STD RANGE MAX MIN
                      SUM USS N;
PROC CHART;
VBAR LIFESPAN/TYPE = PERCENT;
```

To adjust the program to handle another data set, one should change the name of the data set (BATTERY) and the input variable (LIFESPAN) to names appropriate to the new data. The title should be changed also.

The output of this program is as follows:

```
                                ANALYSIS
                                   OF
                              BATTERY DATA
VARIABLE    MEAN      VARIANCE    STANDARD    RANGE    MAXIMUM    MINIMUM
                                  DEVIATION             VALUE      VALUE
LIFESPAN   1274.14   1505155.59   1226.85    4271.00   4285.00     14.00
VARIABLE     SUM        UNCORRECTED          N
                            SS
LIFESPAN   63707.00    154924261.00          50

                                ANALYSIS
                                   OF
                              BATTERY DATA

                           PERCENTAGE BAR CHART
      PERCENTAGE
            !      *****
            !      *****
         50 +      *****
            !      *****
            !      *****
            !      *****
            !      *****
         40 +      *****
            !      *****
            !      *****
            !      *****
            !      *****
         30 +      *****
            !      *****
            !      *****
            !      *****
            !      *****
         20 +      *****   *****
            !      *****   *****
            !      *****   *****
            !      *****   *****
            !      *****   *****                           *****
         10 +      *****   *****                           *****
            !      *****   *****                           *****
            !      *****   *****   *****                   *****
            !      *****   *****   *****   *****   *****   *****
            !      *****   *****   *****   *****   *****   *****
            ----------------------------------------------------
                    600     1200    1800    2400    3000    3600
                             LIFESPAN MIDPOINT
```

CHAPTER 7

ESTIMATION

In Chap. 6, we found that once the family to which a random variable belongs is determined, the problem of approximating or *estimating* the numerical value of pertinent parameters remains. Even though we were able to define sample statistics that allow us to estimate the mean, variance, and standard deviation of a random variable in a logical manner, we were unable to assess their effectiveness. In this chapter we consider the mathematical properties of these statistics. We also present a brief introduction to the theory of estimation. The ideas developed here will be used extensively throughout the remainder of the text.

7.1 POINT ESTIMATION

In an estimation problem there is at least one parameter θ whose value is to be approximated on the basis of a sample. The approximation is done by using an appropriate statistic. A statistic used to approximate or estimate a population parameter θ is called a *point estimator* for θ and is denoted by $\hat{\theta}$ (the symbol $\hat{}$ is called a "hat"); the numerical value assumed by this statistic when evaluated for a given sample is called a point *estimate* for θ. For example, in estimating the mean coal consumption by electric utilities for a given year (see Example 6.3.1), the statistic $\bar{X}$ was used. Thus, $\bar{X}$ is a point estimator for μ and we write $\hat{\mu} = \bar{X}$. In Example 6.3.1, we evaluated this statistic for a particular sample and obtained the value 408.3 million tons. This number is called a point estimate for μ. Note that there is a difference in the terms *estimator* and *estimate*. The estimator is the statistic used to generate the estimate; it is a random variable. An estimate is a number.

Once a logical point estimator for a parameter θ has been developed, the natural question to ask is: "How good is this estimator?" Obviously, we want the estimator to generate estimates that can be expected to be close in value to θ. This can be expected to occur if the estimator $\hat{\theta}$ possesses two properties. In particular, we would like

1. $\hat{\theta}$ to be *unbiased* for θ.
2. $\hat{\theta}$ to have small variance for large sample sizes.

The word *unbiased* is a technical term whose definition is given here.

Definition 7.1.1 (Unbiased). An estimator $\hat{\theta}$ is an unbiased estimator for a parameter θ if and only if $E[\hat{\theta}] = \theta$.

Recall that $\hat{\theta}$ is a statistic; therefore, it is also a random variable and, as such, has a mean, or expected, value. To say that $\hat{\theta}$ is unbiased for θ implies that the mean of the estimator $\hat{\theta}$ is equal to the parameter θ that it is estimating. Thus an estimator $\hat{\mu}$ is an unbiased estimator for μ if and only if $E[\hat{\mu}] = \mu$; an estimator $\hat{\sigma}^2$ is unbiased for σ^2 if and only if $E[\hat{\sigma}^2] = \sigma^2$; an estimator $\hat{\sigma}$ is unbiased for σ if and only if $E[\hat{\sigma}] = \sigma$. Let us reexamine the estimators $\overline{X}$, S^2, and S developed in Chap. 6 in light of this new definition.

Theorem 7.1.1. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from a distribution with mean μ. The sample mean, $\overline{X}$, is an unbiased estimator for μ.

Proof. By Definition 6.3.1

$$E[\overline{X}] = E[1/n(X_1 + X_2 + X_3 + \cdots + X_n)]$$

By the Rules for Expectation (Theorem 3.3.1)

$$E[\overline{X}] = 1/n(E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_n])$$

Since $X_1, X_2, X_3, \ldots, X_n$ constitutes a random sample from a distribution with mean μ, each of these random variables has mean μ. Therefore

$$E[\overline{X}] = 1/n\Big(\underbrace{\mu + \mu + \mu + \cdots + \mu}_{n \text{ terms}}\Big) = 1/n(n\mu) = \mu$$

and the proof is complete.

It is important to realize that since $\hat{\theta}$ is a statistic, in repeated sampling the estimates generated will vary from sample to sample. To say that $\hat{\theta}$ is unbiased for θ implies that these estimates vary about θ; it also implies that the *average* value of these estimates can be expected to lie reasonably close to θ. For example, since $\bar{X}$ is unbiased for μ, for k repetitions of an experiment the observed sample means $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k$ will vary about μ and the *average* value of these k estimates should lie reasonably close to μ.

It is equally important to understand what the term unbiased does *not* imply. It does not imply that any *one* estimate will be close in value to the parameter being estimated. In reference to Example 6.3.1, the estimated mean coal consumption by electric utilities was $\hat{\mu} = \bar{x} \doteq 408.3$ million tons. This estimate is unbiased in the sense that it was generated by means of the unbiased estimator $\bar{X}$. This *alone* does not guarantee that the actual mean coal consumption by electric utilities across the country is anywhere close to 408.3 million tons. This is unfortunate. Usually, statistical studies are not repeated over and over so that the estimates obtained can be averaged. Usually only one sample is drawn; one estimate is obtained. To have some assurance that this estimate is close in value to θ, the parameter being estimated, ideally the estimator used not only should be unbiased, but also should have small variance for large sample sizes. In this way, even though the estimated values fluctuate about θ, the variability is small. Each estimate produced can be expected to be fairly close in value to θ. Theorem 7.1.2 shows that $\bar{X}$ has this property.

Theorem 7.1.2. Let $\bar{X}$ be the sample mean based on a random sample of size n from a distribution with mean μ and variance σ^2. Then

$$\operatorname{Var} \bar{X} = \frac{\sigma^2}{n}$$

The proof of this theorem is based on the Rules for Variance (Theorem 3.3.3) and is similar to that of Theorem 7.1.1. Note that since σ^2 is constant, as the sample size n increases the variance of $\bar{X}$, σ^2/n, decreases and can be made as small as we wish by choosing n sufficiently large. This implies that a sample mean based on a large sample can be expected to lie reasonably close to μ; one based on a small sample may vary widely from the actual population mean. This points out the advantages of working with a large sample and the danger of placing too much emphasis on conclusions drawn from small samples. Keep in mind that many of the examples and exercises presented in this text are based on small samples. This is done for illustrative purposes only. It is *not* meant to imply that samples this small are common in research.

In Chap. 6, we defined the sample variance S^2 by dividing $\sum_{i=1}^{n}(X_i - \bar{X})^2$ by $n - 1$. This was done so that the resulting estimator would be unbiased for σ^2. We now prove that this is the case.

Theorem 7.1.3. Let S^2 be the sample variance based on a random sample of size n from a distribution with mean μ and variance σ^2. S^2 is an unbiased estimator for σ^2.

Proof. By definition

$$
\begin{aligned}
E[S^2] &= E\left[\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}\right] \\
&= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (X_i - \mu + \mu - \overline{X})^2\right] \\
&= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^{n} (X_i - \mu) + \sum_{i=1}^{n} (\overline{X} - \mu)^2\right] \\
&= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X} - \mu)\frac{n\left(\sum X_i - n\mu\right)}{n} + n(\overline{X} - \mu)^2\right] \\
&= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (X_i - \mu)^2 - 2n(\overline{X} - \mu)^2 + n(\overline{X} - \mu)^2\right] \\
&= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2\right] \\
&= \frac{1}{n-1} \left[\sum_{i=1}^{n} E\left[(X_i - \mu)^2\right] - nE\left[(\overline{X} - \mu)^2\right]\right]
\end{aligned}
$$

Note that since $X_1, X_2, X_3, \ldots, X_n$ is a random sample from a distribution with variance σ^2, $E[(X_i - \mu)^2] = \sigma^2$ for each $i = 1, 2, 3, \ldots, n$. Note that by Theorems 7.1.2 and 7.1.1 $\operatorname{Var} \overline{X} = E[(\overline{X} - \mu)^2] = \sigma^2/n$. By substitution, we obtain

$$
\begin{aligned}
E[S^2] &= \frac{1}{n-1}\left[\sum_{i=1}^{n} \sigma^2 - n\sigma^2/n\right] \\
&= \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \sigma^2
\end{aligned}
$$

and the proof is complete.

It should be noted that even though S^2 is an unbiased estimator for σ^2, it can be shown that S is not unbiased for σ (see Exercise 8). This emphasizes the fact that unbiasedness is desirable in an estimator but not essential.

7.2 THE METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD

In this section we consider two methods for deriving point estimators for distribution parameters. The first, called the *method of moments*, is a simple method that was first proposed by Karl Pearson in 1894. The second, called the *method of maximum likelihood*, is more complex. It was used by C. F. Gauss to solve isolated problems over 170 years ago. In the early 1900s the method was formalized by R. A. Fisher and has been used extensively since that time.

To begin, we note that terms of the form $E[X^k]$ $(k = 1, 2, 3, \ldots)$ are called the *kth moments* for X. Since an expectation is a theoretical average, logic implies that the moments for X can be estimated via an arithmetic average. That is, an estimator M_k for $E[X^k]$ based on a random sample of size n is

$$M_k = \sum_{i=1}^{n} \frac{X_i^k}{n}$$

The method of moments exploits the fact that in many cases the moments for X can be expressed as a function of θ, the parameter to be estimated. We can often obtain a reasonable estimator for θ by replacing the theoretical moments by their estimators and solving the resulting equation for $\hat{\theta}$.

You have already used the technique quite naturally in solving some of the problems in the last section! We now formalize the idea. The technique is illustrated by finding the method of moments estimator for the parameter p of a binomial random variable.

Example 7.2.1. A forester plants five rows of twenty pine seedlings each to serve as eventual windbreaks. The soil and wind conditions to which the seedlings are subjected are identical. The variable being studied is X, the number of seedlings per row that survive the first winter. We are dealing with a random sample of size $m = 5$ from a binomial distribution with parameters $n = 20$ and p unknown. We want to use the method of moments to derive an estimator for p. To do so, note that since X is binomial,

$$E[X] = np = 20p$$

We now replace the first moment of X, $E[X]$, by its estimator $M_1 = (\sum_{i=1}^{5} X_i)/5 = \bar{X}$ to obtain the equation

$$\bar{X} = 20\hat{p}$$

This equation is solved for $\hat{p}$ to obtain the estimator

$$\hat{p} = \bar{X}/20$$

When the experiment is conducted, these data result:

$$x_1 = 18 \qquad x_3 = 15 \qquad x_5 = 20$$

$$x_2 = 17 \qquad x_4 = 19$$

For these data, $\bar{x} = (\sum_{i=1}^{5} x_i)/5 = 17.8$. The method of moments estimate for p, the probability that a seedling with survive the first winter is

$$\hat{p} = \bar{x}/20 = 17.8/20 = .89$$

Occasionally there are two parameters θ_1 and θ_2 to be estimated from a single sample. To use the method of moments in this case, we must obtain two equations relating the moments of the distribution to these parameters. We then replace the theoretical moments by their estimators and solve the resulting equations simultaneously for $\hat{\theta}_1$ and $\hat{\theta}_2$. This idea is illustrated by finding estimators for α and β, the parameters that identify the gamma distribution.

Example 7.2.2. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a gamma distribution with parameters α and β. From Theorem 4.3.2, we know that $E[X] = \alpha\beta$ and $\text{Var } X = \alpha\beta^2$. Recall that since $\text{Var } X = E[X^2] - (E[X])^2$, the first two moments of X are functions of α and β. The equations relating the moments to these unknown parameters are

$$E[X] = \alpha\beta$$

$$E[X^2] - (E[X])^2 = \alpha\beta^2$$

We now replace $E[X]$ and $E[X^2]$ by their estimators M_1 and M_2 respectively to obtain

$$M_1 = \hat{\alpha}\hat{\beta}$$

$$M_2 - M_1^2 = \hat{\alpha}\hat{\beta}^2$$

Solving this set of equations simultaneously, we see that

$$M_2 - M_1^2 = M_1\hat{\beta}$$

This implies that

$$\hat{\beta} = (M_2 - M_1^2)/M_1$$

and

$$\hat{\alpha} = M_1/\hat{\beta} = M_1^2/(M_2 - M_1^2)$$

Maximum Likelihood Estimators

The maximum likelihood method for deriving estimators is more complex than the method of moments. However, it is based on an appealing notion. Recall that the density f for a random variable X usually has at least one parameter θ associated with it. Assume that we have a random sample $x_1, x_2, x_3, \ldots, x_n$ available. The method of maximum likelihood in a sense picks out of all the possible values of θ the one most likely to have produced these observations. Before formalizing the method, let us demonstrate the idea in a simple context.

Example 7.2.3. Water samples of a specified size are taken from a river suspected of having been polluted by improper treatment procedures at an upstream sewage disposal plant. Let X denote the number of coliform organisms found per sample and assume that X is a Poisson random variable with parameter k. Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from the distribution of X. We want to determine the value of k that gives the highest probability of observing this sample. Since random sampling implies independence,

$$\begin{aligned} P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] &= P[X_1 = x_1]P[X_2 = x_2] \cdots P[X_n = x_n] \\ &= \prod_{i=1}^{n} P[X_i = x_i] \end{aligned}$$

Recall that the density for X is given by

$$P[X = x] = f(x) = \frac{e^{-k}k^x}{x!} \qquad x = 0, 1, 2, \ldots$$

Therefore the probability of obtaining the given sample is

$$\prod_{i=1}^{n} P[X_i = x_i] = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{e^{-k}k^{x_i}}{x_i!}$$

Note that this probability is a function of k which we denote by $L(k)$. Using the laws of exponents

$$L(k) = \frac{e^{-nk}k^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

This function is called the "likelihood function." It gives us the probability of observing the values $x_1, x_2, \ldots, x_n$ as a function of the parameter k. We want to find the value of k that maximizes this probability. That is, of all the possible values for k, we want to find the one that gives us the highest probability of observing the values that we did observe. To find this value of k we use elementary calculus to maximize the likelihood function. This can be done directly. However, to simplify the process we first take the natural logarithm of $L(k)$ and use the laws of logarithms to simplify the resulting expression

$$\ln L(k) = -nk + \sum_{i=1}^{n} x_i \ln k - \ln \prod_{i=1}^{n} x_i!$$

The value of k that maximizes $\ln L(k)$ also maximizes $L(k)$. Therefore, to complete the derivation we differentiate $\ln L(k)$ with respect to k, set the derivative equal to 0, and solve for k

$$\frac{d \ln L(k)}{dk} = -n + \left(\sum_{i=1}^{n} x_i\right) \Big/ k$$

$$0 = -n + \left(\sum_{i=1}^{n} x_i\right) \Big/ k$$

$$k = \left(\sum_{i=1}^{n} x_i\right) \Big/ n = \bar{x}$$

Since this procedure does not give us the exact value of k but rather provides a logical method for estimating k, we write $\hat{k} = \bar{X}$. That is, the sample mean is the "maximum likelihood estimator" for the parameter k of a Poisson random variable.

Suppose that a random sample of size 4 yields these data:

$$x_1 = 12 \qquad x_2 = 15 \qquad x_3 = 16 \qquad x_4 = 17$$

Since the value of k that is most likely to have produced this sample is $\bar{x} = 15$, it is natural to take this value as our estimate for k.

Although our example involves a discrete random variable, the same general method is used in the continuous case. This method is summarized as follows:

1. Obtain a random sample $x_1, x_2, x_3, \ldots, x_n$ from the distribution of a random variable X with density f and associated parameter θ.
2. Define a function $L(\theta)$ by

$$L(\theta) = \prod_{i=1}^{n} f(x_i)$$

   This function is called the *likelihood function* for the sample.
3. Find the expression for θ that maximizes the likelihood function. This can be done directly or by maximizing $\ln L(\theta)$.
4. Replace θ by $\hat{\theta}$ to obtain an expression for the maximum likelihood estimator for θ.
5. Find the observed value of this estimator for a given sample.

As with the method of moments, the maximum likelihood procedure can be applied when the density for X is characterized by two parameters. We illustrate the technique by finding the maximum likelihood estimators for μ and σ^2, the mean and variance of a normal random variable.

Example 7.2.4. Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from a normal distribution with mean μ and variance σ^2. The density for X is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2}$$

The likelihood function for the sample is a function of both μ and σ. In particular,

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)[(x_i-\mu)/\sigma]^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

The logarithm of the likelihood function is

$$\ln L(\mu, \sigma) = -n \ln\sqrt{2\pi} - n \ln\sigma - (1/2\sigma^2) \sum_{i=1}^{n} (x_i - \mu)^2$$

To maximize this function, we take the partial derivatives with respect to μ and σ;

set these derivatives equal to 0, and solve the equations simultaneously for μ and σ.

$$\begin{cases} \dfrac{\partial \ln L(\mu,\sigma)}{\partial \mu} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \\[2ex] \dfrac{\partial \ln L(\mu,\sigma)}{\partial \sigma} = \dfrac{-n}{\sigma} + \sum_{i=1}^{n}\dfrac{(x_i-\mu)^2}{\sigma^3} = \dfrac{-n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3} \end{cases}$$

$$\begin{cases} \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \\[2ex] \dfrac{-n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3} = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n} x_i - n\mu = 0 \quad \text{or} \quad \mu = \left(\sum_{i=1}^{n} x_i\right)\Big/ n = \bar{x} \\[2ex] -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0 \quad \text{or} \quad \sigma^2 = \left[\sum_{i=1}^{n}(x_i-\mu)^2\right]\Big/ n \end{cases}$$

Realizing that these are not the true values of μ and σ^2 but are only estimates, we see that the maximum likelihood estimators for these parameters are

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]\Big/ n$$

The method of moments estimator for a parameter and the maximum likelihood estimator often agree. However, if they do not the maximum likelihood estimator is usually preferred.

7.3 FUNCTIONS OF RANDOM VARIABLES—DISTRIBUTION OF $\bar{X}$

There is one drawback to point estimation. It yields a single value for the unknown parameter θ. Is there any assurance that this estimate is even close in value to θ? The best answer is that in most cases, the point estimators used are logical. To get an idea not only of the value of the parameter being estimated, but also of the accuracy of the estimate, researchers turn to the method of *interval estimation* or *confidence intervals*. An interval estimator is what the name implies. It is a random interval, an interval whose endpoints L_1 and L_2 are each statistics. It is used to determine a numerical interval based on a sample. It is hoped that the numerical interval obtained will contain the population parameter being estimated. By expanding from a point to an interval, we create a little room for error and in so doing gain the ability, based on probability theory, to report the confidence that we have in the estimate.

In later chapters we shall derive confidence intervals for many important parameters. To do so we must know the distribution of some key random variables. In this section we consider a technique for identifying the distribution of a random variable from its moment generating function. This technique depends on the result given in Theorem 7.3.1.

Theorem 7.3.1. Let X and Y be random variables with moment generating functions $m_X(t)$ and $m_Y(t)$ respectively. If $m_X(t) = m_Y(t)$ for all t in some open interval about 0, then X and Y have the same distribution.

The proof of this theorem is based on transform theory and is beyond the scope of this text. The theorem implies that the moment generating function, when it exists, serves as a "fingerprint" for the random variable. We illustrate this idea by proving Theorem 4.4.3, the "standardization" theorem for normal random variables.

Example 7.3.1. Let X be a normal random variable with mean μ and variance σ^2. Recall from Theorem 4.4.1, that the moment generating function for X is

$$m_X(t) = E[e^{tX}] = e^{\mu t + \sigma^2 t^2/2}$$

The moment generating function for a standard normal random variable Z is

$$m_Z(t) = e^{0t + (1)^2 t^2/2} = e^{t^2/2}$$

Let $Y = (X - \mu)/\sigma = (1/\sigma)X - \mu/\sigma$. The moment generating function for Y is given by

$$\begin{aligned} m_Y(t) = E[e^{tY}] &= E[e^{(t/\sigma)X - (\mu/\sigma)t}] \\ &= E[e^{(t/\sigma)X} e^{(-\mu/\sigma)t}] \\ &= e^{(-\mu/\sigma)t} E[e^{(t/\sigma)X}] \end{aligned}$$

Note that $E[e^{(t/\sigma)X}] = m_X(t/\sigma) = e^{\mu(t/\sigma) + \sigma^2 t^2/2\sigma^2}$. Substituting,

$$m_Y(t) = e^{(-\mu/\sigma)t} e^{(\mu/\sigma)t + t^2/2} = e^{t^2/2} = m_Z(t)$$

We have shown that Y and Z have the same moment generating function. By Theorem 7.3.1, these variables have the same distribution. In particular, they are both *standard normal* random variables.

Many of the statistics used in data analysis entail summing a collection of random variables. The following theorem together with Theorem 7.3.1 will help determine the distribution of such statistics.

Theorem 7.3.2. Let X_1 and X_2 be independent random variables with moment generating functions $m_{X_1}(t)$ and $m_{X_2}(t)$ respectively. Let $Y = X_1 + X_2$. The moment generating function for Y is given by

$$m_Y(t) = m_{X_1}(t) m_{X_2}(t)$$

Proof. By definition

$$m_Y(t) = E[e^{tY}] = E[e^{tX_1 + tX_2}] = E[e^{tX_1}e^{tX_2}]$$

Since X_1 and X_2 are independent, e^{tX_1} and e^{tX_2} are also independent. By Theorem 5.2.2

$$m_Y(t) = E[e^{tX_1}e^{tX_2}] = E[e^{tX_1}]E[e^{tX_2}] = m_{X_1}(t)m_{X_2}(t)$$

This theorem can be extended easily to include a sum of more than two random variables. That is, we can say that the moment generating function for the sum of a finite number of *independent* random variables is the product of the moment generating functions of the individual variables. The requirement that the random variables be independent is not restrictive since in most cases the sum of interest is a function of the elements of a random sample. The term "random sample" implies independence. (See Definition 6.1.1.) Theorem 7.3.2 is illustrated by showing that the sum of a collection of independent normal random variables is normal.

Example 7.3.2. Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots, \sigma_n^2$ respectively. Let $Y = X_1 + X_2 + X_3 + \cdots + X_n$. Note that the moment generating function for X_i is given by

$$m_{X_i}(t) = e^{\mu_i t + \sigma_i^2 t^2/2} \qquad i = 1, 2, 3, \ldots, n$$

and the moment generating function for Y is

$$m_Y(t) = \prod_{i=1}^{n} m_{X_i}(t) = \exp\left[\left(\sum_{i=1}^{n} \mu_i\right)t + \left(\sum_{i=1}^{n} \sigma_i^2\right)t^2 \Big/ 2\right]$$

The function on the right is the moment generating function for a normal random variable with mean $\mu = \sum_{i=1}^{n}\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n}\sigma_i^2$.

Distribution of $\bar{X}$

One of the more useful statistics that we have studied is $\bar{X}$, the sample mean. Since $\bar{X}$ is a statistic it is also a random variable. It makes sense to ask: "What is the distribution of $\bar{X}$?" We have already seen that the center of location for $\bar{X}$ is μ, the mean of the population from which the sample is drawn. We have also seen that its variance is σ^2/n, the original population variance divided by the sample size. We have not yet mentioned the type of distribution possessed by this statistic. Does $\bar{X}$ follow some distribution such as the gamma, uniform, or normal that we have already studied or must we introduce a new distribution now? The next theorem, whose derivation is outlined in Exercise 32, will help us answer this question.

Theorem 7.3.3. Let X be a random variable with moment generating function $m_X(t)$. Let $Y = \alpha + \beta X$. The moment generating function for Y is

$$m_Y(t) = e^{\alpha t}m_X(\beta t)$$

We illustrate the use of this theorem in a numerical context.

Example 7.3.3. Let X denote the maximum wind speed per day recorded at the weather station of a particular locality. Assume that X is normally distributed with mean 10 mph and standard deviation 4 mph. Engineers are constructing a bridge over a deep canyon in the area. They suspect that the maximum wind speed at the bridge site is given by $Y = 2X - 5$. What is the distribution of Y? To answer this question we first note that the moment generating function for X is

$$m_X(t) = e^{\mu t + \sigma^2 t^2/2}$$
$$= e^{10t + 16t^2/2}$$

We next apply Theorem 7.3.3 with $\alpha = -5$ and $\beta = 2$ to see that the moment generating function for Y is

$$m_Y(t) = e^{-5t} e^{10(2t) + 16(2t)^2/2}$$
$$= e^{15t + 64t^2/2}$$

This is the moment generating function for a normal random variable with mean 15 mph and variance 64. Since the moment generating function for a random variable is its fingerprint, we know that the maximum speed at the bridge site is normally distributed with an average speed of 15 mph and a standard deviation of 8 mph.

Theorem 7.3.3 is interesting in its own right but its primary purpose at this time is to help us derive the next very important theorem. This theorem answers the question posed earlier concerning the distribution of $\overline{X}$. In particular, it assures us that when sampling from a *normal* distribution the random variable $\overline{X}$ will itself be *normally* distributed.

Theorem 7.3.4 (Distribution of $\overline{X}$—normal population). Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then $\overline{X}$ is normally distributed with mean μ and variance σ^2/n.

The derivation of this theorem is not hard. It is outlined in Exercises 33 to 36. We feel that by working through the derivation for yourself you will have a better understanding of the point being made. The other exercises presented are also important. They contain some results that will have major practical consequences later. Be sure to given them all a try!

7.4 INTERVAL ESTIMATION AND THE CENTRAL LIMIT THEOREM

As mentioned previously, point estimation does not give us the ability to report the accuracy of our estimate. To do this, we must turn to the method of interval estimation. The statistics used to extend a point estimate for a parameter θ to an interval of values that should contain the true value of θ vary from parameter to parameter. However, the method for deriving these statistics is basically the same in each case. In this section we illustrate the method by deriving a "confidence interval" for the mean of a normal random variable when its variance is assumed to be known. In later chapters we apply the general technique illustrated here to find confidence intervals for other important parameters.

The term "confidence interval" is a technical term which we now define.

Definition 7.4.1 (Confidence interval). A $100(1-\alpha)\%$ confidence interval for a parameter θ is a random interval $[L_1, L_2]$ such that

$$P[L_1 \le \theta \le L_2] \doteq 1-\alpha$$

regardless of the value of θ.

One general statement will guide in the construction of most of the confidence intervals presented in this text.

> To construct a $100(1-\alpha)\%$ confidence interval for a parameter θ we shall find a random variable whose expression involves θ and whose probability distribution is known at least approximately.

To use this guideline to find a $100(1-\alpha)\%$ confidence interval for the mean of a normal random variable whose variance is known, we must find a random variable whose expression involves μ and whose distribution is known. This is easy to do. Note that in Theorem 7.3.4, we showed that under the given conditions the sample mean, $\bar{X}$, is normally distributed with mean μ and variance σ^2/n. This implies that the random variable

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

is *standard normal*. Note that this random variable involves the parameter μ and its distribution is known. We illustrate how this random variable can be used to generate a 95% confidence interval for μ. The technique used can be generalized easily to obtain any desired degree of confidence.

Example 7.4.1. Acute myeloblastic leukemia is among the most deadly of cancers. Past experience indicates that the time in months that a patient survives after initial diagnosis of the disease is normally distributed with a mean of 13 months and a standard deviation of three months. A new treatment is being investigated which should prolong the average survival time without affecting variability. Let $X_1, X_2, X_3, \ldots, X_n$ denote a random sample from the distribution of X, the survival time under the new treatment. We are assuming that X is normally distributed with $\sigma^2 = 9$ and μ unknown. We want to find statistics L_1 and L_2 so that $P[L_1 \le \mu \le L_2] \doteq .95$. To do so, consider the partition of the standard normal curve shown in Fig. 7.1. It can be seen that

$$P[-1.96 \le Z \le 1.96] = .95$$

In this case, $Z = (\bar{X}-\mu)/(\sigma/\sqrt{n})$, and hence we may conclude that

$$P\left[-1.96 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1.96\right] = .95$$

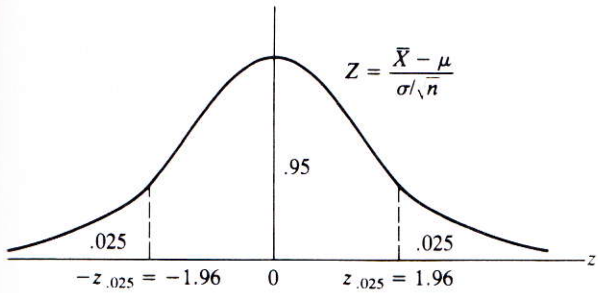


FIGURE 7.1
Partition of Z to obtain a 95% confidence interval for μ.

To find L_1 and L_2, we algebraically isolate μ in the center of the preceding inequality as follows:

$$P\left[-1.96\sigma/\sqrt{n} \le \overline{X} - \mu \le 1.96\sigma/\sqrt{n}\right] = .95$$

$$P\left[-\overline{X} - 1.96\sigma/\sqrt{n} \le -\mu \le -\overline{X} + 1.96\sigma/\sqrt{n}\right] = .95$$

$$P\left[\overline{X} - 1.96\sigma/\sqrt{n} \le \mu \le \overline{X} + 1.96\sigma/\sqrt{n}\right] = .95$$

From this we see that the lower and upper bounds for a 95% confidence interval are

$$L_1 = \overline{X} - 1.96\sigma/\sqrt{n} \qquad L_2 = \overline{X} + 1.96\sigma/\sqrt{n}$$

These statistics have the property that in repeated sampling from the population, 95% of the numerical intervals generated are expected to contain μ; by chance, 5% will not. This idea is illustrated in Fig. 7.2.

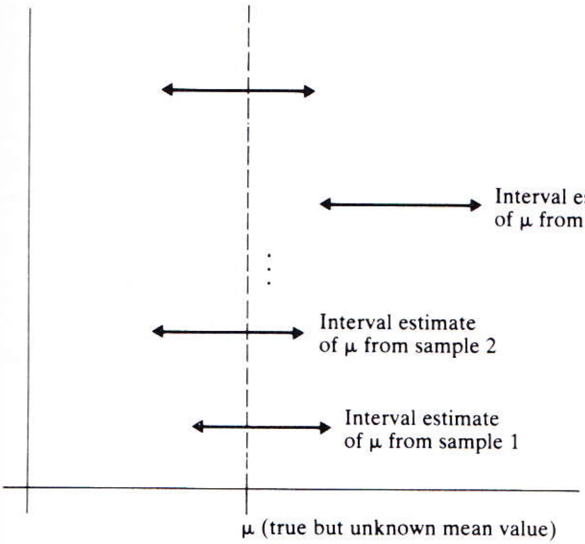


FIGURE 7.2
Of the intervals constructed by using $[L_1, L_2]$, 95% are expected to contain μ, the true but unknown population mean.

Note that since we are assuming that σ^2 is known, the confidence bounds, $\overline{X} \pm 1.96\sigma/\sqrt{n}$, just derived are *statistics*. Given a particular set of observations on X, their numerical values can be determined easily as demonstrated in Example 7.4.2.

Example 7.4.2. When the experiment of Example 7.4.1 is conducted these observations on X, the survival time under the new treatment, result:

8.0	13.6	13.2	13.6
12.5	14.2	14.9	14.5
13.4	8.6	11.5	16.0
14.2	19.0	17.9	17.0

Based on these data, $\hat{\mu} = \bar{x} = 13.88$ months. This point estimate is extended to a 95% confidence interval by evaluating the statistics L_1 and L_2. In particular

$$\begin{aligned} L_1 = \bar{x} - 1.96\sigma/\sqrt{n} &= 13.88 - 1.96(3/\sqrt{16}) \\ &= 13.88 - 1.47 \\ &= 12.41 \text{ months} \\ L_2 = \bar{x} + 1.96\sigma/\sqrt{n} &= 13.88 + 1.47 \\ &= 15.35 \text{ months} \end{aligned}$$

Based on these data, the interval estimate for μ is [12.41, 15.35]. Does the true mean survival time for patients receiving the new treatment really lie between 12.41 and 15.35 months? Unfortunately, there is no way of knowing. The interval [12.41, 15.35] is a 95% confidence interval. This means that the procedure used is expected to trap μ 95% of the time. We hope that the interval obtained from our particular sample does so.

To obtain the general formula for a $100(1 - \alpha)\%$ confidence interval on the mean of a normal random variable whose variance is known, we need only partition the standard normal curve as shown in Fig. 7.3. The algebraic argument of Example 7.4.1 goes through exactly as presented with the point $z_{.025} = 1.96$

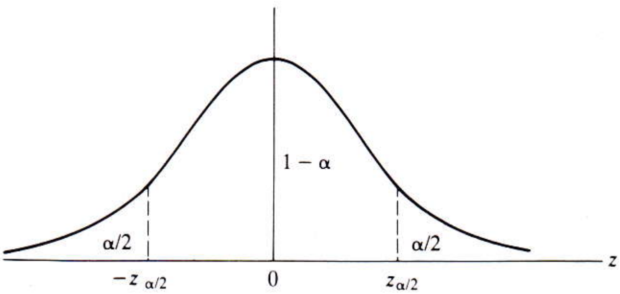


FIGURE 7.3
Partition of Z to obtain a $100(1 - \alpha)\%$ confidence interval for μ.

being replaced by $z_{\alpha/2}$. This change results in the general formula given in Theorem 7.4.1.

Theorem 7.4.1 ($100(1 - \alpha)\%$ Confidence interval on μ when σ^2 is known). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. A $100(1 - \alpha)\%$ confidence interval on μ is given by

$$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

Central Limit Theorem

There is one further point to be made. Theorem 7.4.1 does require that the base variable X be normal. If this condition is not satisfied, then the confidence bounds given can be used as long as the sample is not too small. Empirical studies have shown that for samples as small as 25, the above bounds are usually satisfactory even though approximate. This is due to a remarkable theorem, first formulated in the early nineteenth century by Laplace and Gauss. This theorem, known as the *Central Limit Theorem*, gives the distribution of $\bar{X}$ when sampling from a distribution that is not necessarily normal.

Theorem 7.4.2 (Central Limit Theorem). Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, $\bar{X}$ is approximately normal with mean μ and variance σ^2/n. Furthermore, for large n, the random variable $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is approximately standard normal.

Please note the differences between the Central Limit Theorem and Theorem 7.3.4. The former does not require that sampling be from a normal distribution whereas normality is assumed in the latter; the former claims that $\bar{X}$ will be approximately normally distributed for large sample sizes whereas the latter claims that $\bar{X}$ will be exactly normally distributed regardless of the sample size involved.

The Central Limit Theorem is important to us for two reasons. First it allows us to make inferences on the mean of a distribution based on relatively large samples without having to be overly concerned as to whether or not we are sampling from a normal distribution; second, it allows us to justify analytically the normal approximations to discrete distributions presented in Chap. 4. Recall that in that chapter we considered the normal approximation to the binomial distribution. (Theorem 4.6.1.) At that time, the approximation was justified graphically. We can now give a more rigorous argument to justify this approximation procedure.

Example 7.4.3. Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from a point binomial distribution (see Exercise 43, Chap. 3). Recall that each of these random variables is binomial with parameters 1 and p. Each has mean p, variance $p(1 - p)$ and moment generating function of the form $q + pe^t$. Let $X = \sum_{i=1}^{n} X_i$. Since $X_1, X_2, \ldots, X_n$ are independent, the moment generating function for X is given by

$$m_X(t) = \prod_{i=1}^{n} (q + pe^t) = (q + pe^t)^n$$

This is the moment generating function for a binomial random variable with parameters n and p. By the Central Limit Theorem, $\bar{X} = (\sum_{i=1}^{n} X_i)/n = X/n$ is approximately normal with mean p and variance $p(1-p)/n$. Now consider the binomial random variable $n(X/n) = X$. Since X is a linear function of the approximately normal random variable X/n, we can apply Exercise 35 with $a_1 = n$ and $a_i = 0$, $i \neq 1$, to conclude that X is approximately normal with mean np and variance $[n^2 p(1-p)]/n = np(1-p)$.

Exercises 43, 44, 49, 50, 55, and 56 will give you practice in the application of the Central Limit Theorem.

CHAPTER SUMMARY

In this chapter we considered the ideas of point and interval estimation. We introduced three types of point estimators. These are unbiased estimators, method of moments estimators, and maximum likelihood estimators. Unbiased estimators are estimators whose mean value is equal to the parameter being estimated. We showed that $\bar{X}$ is unbiased for μ, that S^2 is unbiased for σ^2, but that S is not unbiased for σ. Method of moments estimators are derived by noting that the parameters that characterize a distribution are often functions of the kth moments of the distribution. Maximum likelihood estimators are found by choosing the value of the parameter θ that maximizes the likelihood function. In this way, in some sense we pick out of all possible values of θ the one that is most likely to have produced the observed data.

In order to develop the idea of interval estimation, we introduced some theorems that help us determine the distribution of a random variable. In particular, we noted that the moment generating function for a random variable is its "fingerprint." To determine its distribution we look at its moment generating function. This technique was used to verify the standardization theorem used in earlier chapters. It was also used to show that a linear function of independent normal random variables is normal, that a sum of independent chi-squared random variables is chi-squared and that $\bar{X}$ is normally distributed when sampling from a normal distribution. (See Exercise 38.)

We introduced the general concept of a $100(1-\alpha)\%$ confidence interval on a parameter θ. This is a random interval, an interval of the form $[L_1, L_2]$, where L_1 and L_2 are statistics with the property that "a priori" θ will be trapped between L_1 and L_2 with probability $1-\alpha$. We used information just developed on the distribution of $\bar{X}$ to develop specific formulas for constructing a $100(1-\alpha)\%$ confidence interval on the mean of a normal distribution. Finally, we considered the central limit theorem. This theorem concerns the approximate distribution of $\bar{X}$ when sampling from a nonnormal distribution. It allows us to make inferences on the mean of any distribution when relatively large samples are available. It also allows us to justify some of the approximation techniques presented earlier in the text.

These new terms were introduced:

Point estimator	Point estimate
Unbiased	Weighted mean
kth moments	Likelihood function
Confidence interval or interval estimator	Interval estimate

EXERCISES

Section 7.1

1. Let $X_1, X_2, X_3, \ldots, X_{20}$ be a random sample from a distribution with mean 8 and variance 5. Find the mean and variance of $\bar{X}$.
2. Let $X_1, X_2, X_3, \ldots, X_{15}$ be a random sample from a Poisson distribution with parameter λs. Given an unbiased estimator for this parameter.
3. Let X denote the number of paint defects found in a square yard section of a car body painted by a robot. These data are obtained:

8	5	0	10
0	3	1	12
2	7	9	6

   Assume that X has a Poisson distribution with parameter λs.
   (*a*) Find an unbiased estimate for λs.
   (*b*) Find an unbiased estimate for the average number of flaws per square yard.
   (*c*) Find an unbiased estimate for the average number of flaws per square foot.
4. An interactive computer system is available at a large installation. Let X denote the number of requests for this system received per hour. Assume that X has a Poisson distribution with parameter λs. These data are obtained:

25	20	20
30	24	15
10	23	4

   (*a*) Find an unbiased estimate for λs.
   (*b*) Find an unbiased estimate for the average number of requests received per hour.
   (*c*) Find an unbiased estimate for the average number of requests received per quarter hour.
5. Let X_1, X_2, X_3, X_4, X_5 be a random sample from a binomial distribution with $n = 10$ and p unknown.
   (*a*) Show that $\bar{X}/10$ is an unbiased estimator for p.
   (*b*) Estimate p based on these data: $3, 4, 4, 5, 6$.
6. An experiment is conducted to study the effect of a power surge on data stored in a digital computer. A "word" is a sequence of eight bits. Each bit is either "on" (activated) or "off" (not activated) at any given time. Twenty 8-bit words are stored and a power surge is induced. Let X denote the number of bit reversals that result

per word. Assume that X is binomially distributed with $n = 8$ and p, the probability of a bit reversal, unknown. These data result:

1	0	0	0
0	0	1	1
0	1	2	1
1	0	1	0
2	2	3	0

(a) Find an unbiased estimate for p.
(b) Based on the estimate for p just found, approximate the probability that in another eight-bit word a similar power surge will result in no bit reversals.
(c) A data line utilizes 64 bits. Based on the estimate for p just found, approximate the probability that at most one bit reversal will occur.

7. Stress tests are conducted on fiberglass rods used in communications networks. The random variable studied is X, the distance in inches from the anchored end of the rod to the crack location when the rod is subjected to extreme stress. Assume that X is uniformly distributed over the interval $(0, b)$. These data are obtained on 10 test rods:

10	7	11	12	8
8	9	10	9	13

(a) Find an unbiased estimate for the average distance from the anchored end of the rod to the crack.
(b) Find an unbiased estimate for the variance of X.
(c) Find an unbiased estimate for b.
(d) Find an estimate for σ, the standard deviation of X. Is this estimate unbiased?

8. Note that S is a statistic and, unless X is constant, its value will vary from sample to sample. Therefore $\operatorname{Var} S > 0$. To show that S is not unbiased for σ use proof by contradiction. That is, assume that $E[S] = \sigma$ and obtain a contradiction. *Hint:* Use Theorem 3.3.2.

9. (*Weighted means.*) Assume that one has k independent random samples of sizes $n_1, n_2, n_3, \ldots, n_k$ from the same distribution. These samples generate k unbiased estimators for the mean, namely, $\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots, \overline{X}_k$.
(a) Show that the arithmetic average of these estimators, $(\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \cdots + \overline{X}_k)/k$, is also unbiased for μ.
(b) Certain mineral elements required by plants are classed as macronutrients. Macronutrients are measured in terms of their percentage of the dry weight of the plant. Proportions of each element vary in different species and in the same species grown under differing conditions. One macronutrient is sulfur. In a study of winter cress, a member of the mustard family, these data, based on three independent random samples, are obtained:

$$\bar{x}_1 = .8 \qquad \bar{x}_2 = .95 \qquad \bar{x}_3 = .7$$
$$n_1 = 9 \qquad n_2 = 3 \qquad n_3 = 200$$

Use the result of part (a) to obtain an unbiased estimate for μ, the mean proportion of sulfur by dry weight in winter cress. By averaging the three values .8, .95, and .7 to obtain the estimate for μ, each sample is being given the equal importance or "weight." Does this seem reasonable in this problem? Explain.

(*c*) To take sample sizes into account a "weighted" mean is used. This estimator, $\hat{\mu}_W$, is given by

$$\hat{\mu}_W = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \cdots + n_k \bar{X}_k}{n_1 + n_2 + \cdots + n_k}$$

Show that $\hat{\mu}_W$ is an unbiased estimator for μ.

(*d*) Use the data of part (*b*) to find the weighted estimate for the mean proportion of sulfur by dry weight in winter cress. Compare your answer to the estimate found in part (*b*).

10. Let X denote the number of heads obtained when a fair coin is tossed four times.

(*a*) What is $E[X]$ and Var X?

(*b*) Perform the experiment of tossing a fair coin four times and recording the number of heads obtained 10 times. You thus obtain a random sample of size 10 from a binomial distribution with $n = 4$ and $p = \frac{1}{2}$.

(*c*) Based on your 10 observations, estimate the mean and variance of X. Compare your answers to those of your classmates. Do the observed values of $\bar{X}$ fluctuate about the theoretical mean of 2? Do the observed values of S^2 fluctuate about the theoretical variance of 1?

(*d*) Average the values of $\bar{X}$ that you have available. Is the average value close to 2? Average the values of S^2 that you have available. Is the average value of S^2 close to 1?

11. Let X denote the number of heads obtained when a fair coin is tossed four times. Perform this experiment three times and record the value of X for each set of four tosses. In this way you obtain a single sample of size 3 from a binomial distribution with $n = 4$ and $p = \frac{1}{2}$.

(*a*) Find the numerical value of $\bar{X}$ for your sample.

(*b*) Repeat the experiment nine more times recording the value of $\bar{X}$ each time.

(*c*) What is $E[\bar{X}]$? Average your 10 values of $\bar{X}$. Is the average value close to the theoretical mean of 2?

(*d*) What is Var $\bar{X}$? Find the value of S^2 for the 10 observations on $\bar{X}$. Does this value lie close to the theoretical value of $1/3$?

***12.** Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Find $E[\sum_{i=1}^{n}(X_i - \bar{X})^2/n]$. *Hint:* Note that $\sum_{i=1}^{n}(X_i - \bar{X})^2/n = (n-1)S^2/n$.

You have shown that the estimator $(n-1)S^2/n$ for σ^2 tends to underestimate σ^2. This is the reason that we usually use S^2 as an estimator for σ^2 even though division by n rather than $n - 1$ is more logical.

Section 7.2

13. Suppose that when the experiment described in Example 7.2.1 is conducted these data result:

$$x_1 = 13 \qquad x_3 = 15 \qquad x_5 = 17$$
$$x_2 = 12 \qquad x_4 = 10$$

Use the method of moments to estimate p, the probability that a seedling will survive the first winter.

14. Let $X_1, X_2, \ldots, X_m$ be a random sample of size m from a binomial distribution with parameters n, assumed to be known, and p. Show that the method of moments estimator for p is $\hat{p} = \bar{X}/n$.

15. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λs. Find the method of moments estimator for λs. Find the method of moments estimator for λ, the parameter underlying the Poisson process under observation.

16. In studying the traffic flow at an intersection a Poisson process with parameter λ is assumed. The basic unit of time assumed is a minute. These data are obtained on X, the number of vehicles arriving at the intersection during a two-minute period:

2	5	1	3	2
3	0	8	2	5

Use these data to estimate λs, the average number of vehicles arriving during a two-minute period, and λ the average number arriving per minute. (Use the results of Exercise 15.)

17. Use the information obtained in Example 7.2.2 to find an estimator for σ^2, the variance of a gamma random variable. Is the estimator obtained unbiased for σ^2? *Hint:* Express M_1 and M_2 as arithmetic averages and compare your result to that of Theorem 6.3.1.

18. An acid solution made by mixing a powder compound with water is used to etch aluminum. The pH of the solution, X, will vary due to slight variations in the amount of water used, the potency of the dry compound, and the pH of the water itself. Assume that X is gamma distributed with α and β unknown. From these data, estimate α, β, μ, and σ^2 using the method of moments:

1.2	2.0	1.6	1.8	1.1
2.5	2.1	2.6	2.2	1.7
1.5	1.7	2.0	3.0	1.8

19. Find the method of moments estimator for the parameter p of a geometric distribution.

20. Let X be normal with mean μ and variance σ^2 both of which are unknown. Find the method of moments estimators for these parameters. Are the estimators obtained unbiased for their respective parameters? Explain.

21. Carbon dioxide is an odorless, colorless gas which constitutes about .035% by volume of the atmosphere. It affects the heat balance by acting as a one-way screen. It lets in the sun's heat to warm the oceans and the land but blocks some of the infrared heat that is radiated from the earth. This reflected heat is absorbed into the lower atmosphere producing a greenhouse effect which causes the earth's surface to become warmer than it would be otherwise. Systematic measurements of CO_2 began in 1957 with Charles D. Keeling began monitoring at Mauna Loa in Hawaii. (*McGraw-Hill Yearbook of Science and Technology*, 1983, p. 68.)

(*a*) Assume that these CO_2 readings (ppm) are obtained:

319	338	337	339	328
325	340	331	341	336
330	330	321	327	337
320	343	350	322	334
326	349	341	338	332
339	335	338	333	334

Construct a stem-and-leaf diagram for these data using 31, 32, 32, 33, 33, 34, 34, 35 as stems. Graph leaves 0–4 on the first of each repeated stem and leaves 5–9 on the other. Is it reasonable to assume that the CO_2 level in the atmosphere is normally distributed? Explain.

(*b*) Estimate μ and σ^2 using the method of moments estimators.

(*c*) Find an unbiased estimate for σ^2.

22. Based on the data of Exercise 16, what is the maximum likelihood estimate for λ, the average number of vehicles arriving at an intersection per minute?

23. Based on the data of Exercise 21, what are the maximum likelihood estimates for the mean and variance of the atmospheric carbon dioxide level?

24. Let $X_1, X_2, X_3, \ldots, X_m$ be a random sample of size m from a binomial distribution with parameters n, assumed to be known, and p. Find the maximum likelihood estimator for p. Does it differ from the method of moments estimator found in Exercise 14?

25. Let W be an exponential random variable with parameter β unknown. Find the maximum likelihood estimator for β based on a sample of size n. Does it differ from the method of moments estimator?

26. A computer center employs consultants to answer users' questions. The center is open from 9 a.m. to 5 p.m. each weekday. Assume that calls arriving at the center constitute a Poisson process with unknown parameter λ calls per hour. To estimate λ, these observations were obtained on X, the number of calls arriving per hour:

8	6	12	15	12
4	9	7	20	10

(*a*) Find the maximum likelihood estimate for λ.

(*b*) Estimate the average time of arrival of the first call of the day. *Hint:* Consider Theorem 4.3.3.

27. A study of the noise level on takeoff of jet airplanes at a particular airport is studied. The random variable is X, the noise level in decibels of the plane as it passes over the first residential area adjacent to the airport. This random variable is assumed to have a gamma distribution with $\alpha = 2$ and β unknown.

(*a*) Find the maximum likelihood estimate for β based on a sample of size n.

(*b*) Use $\hat{\beta}$ to find an estimate for the mean value of X. Is this estimator unbiased for μ?

(*c*) Find the maximum likelihood estimate for β based on these data:

55	65	60	73	80
64	57	75	62	86
69	100	70	82	65
72	67	61	95	52

(*d*) Estimate the average decibel reading of these jets.

28. Computer terminals have a battery pack that maintains the configuration of the terminal. These packs must be replaced occasionally. Let X denote the life span in years of such a battery. Assume that X is exponentially distributed with unknown parameter β. Find the maximum likelihood estimate for β based on these data:

1.7	4.0	1.9	2.0	1.7
2.1	2.7	4.2	1.8	2.2
3.1	1.5	2.4	6.2	7.0
3.6	1.4	5.0	3.8	1.6

29. To estimate the proportion of defective microprocessor chips being produced by a particular maker, samples of five chips are selected at 10 randomly selected times during the day. These chips are inspected and X, the number of defective chips in each batch of size 5, is recorded. Assume that X is binomially distributed with $n = 5$ and p unknown. Use these data to find the maximum likelihood estimate for p:

1	0	1	2	0
0	0	0	1	0

30. A new material is being tested for possible use in the brake shoes of automobiles. These shoes are expected to last for at least 75,000 miles. Fifteen sets of four of these experimental shoes are subjected to accelerated life testing. The random variable X, the number of shoes in each group of four that fail early, is assumed to be binomially distributed with $n = 4$ and p unknown. Find the maximum likelihood estimate for p based on these data:

1	0	1	0	2	1	0	1
0	1	0	0	0	1	0	

If an early failure rate in excess of 10% is unacceptable from a business point of view, would you have some doubts concerning the use of this new material? Explain.

Section 7.3

31. In each part the moment generating function for a random variable X is given. Identify the family to which the random variable belongs and give the numerical values of pertinent distribution parameters.
(a) $m_X(t) = e^{2t+9t^2/2}$
(b) $m_X(t) = e^{8t^2}$
(c) $m_X(t) = .25e^t/(1 - .75e^t)$
(d) $m_X(t) = (.5 + .5e^t)^5$
(e) $m_X(t) = e^{6(e^t-1)}$
(f) $m_X(t) = (1 - 3t)^{-5}$
(g) $m_X(t) = (1 - 2t)^{-8}$
(h) $m_X(t) = (1 - .5t)^{-1}$

32. (a) Let X be a random variable with moment generating function $m_X(t)$. Let $Y = \alpha + \beta X$. Show that $m_Y(t) = e^{\alpha t} m_X(\beta t)$. *Hint:* $m_Y(t) = E[e^{tY}] = E[e^{(\alpha+\beta X)t}]$.
(b) Let X be a normal random variable with mean 10 and variance 4. Find the moment generating function for the random variable $Y = 8 + 3X$. What is the distribution of Y?

33. Let $X_1, X_2, X_3, \ldots, X_n$ be a collection of independent random variables with moment generating functions $m_{X_i}(t)$ ($i = 1, 2, 3, \ldots, n$ respectively). Let $a_0, a_1, a_2, \ldots, a_n$ be real numbers and let

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \cdots + a_n X_n$$

Show that the moment generating function for Y is given by

$$m_Y(t) = e^{a_0 t} \prod_{i=1}^{n} m_{X_i}(a_i t)$$

Note that this extends the result of Exercise 32(a) to more than one variable.

34. Let X_1 and X_2 be independent normal random variables with means 2 and 5 and variances 9 and 1 respectively. Let $Y = 3X_1 + 6X_2 - 8$. Use Exercise 33 to find the moment generating function for Y. What is the distribution of Y?

35. Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means μ_i and σ_i^2 ($i = 1, 2, 3, \ldots, n$ respectively). Let $a_0, a_1, a_2, \ldots, a_n$ be real numbers and let

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

Use Exercise 33 to show that Y is normal with mean $\mu = a_0 + \sum_{i=1}^{n} a_i \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$.

36. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Use Exercise 35 to show that $\overline{X}$ is normal with mean μ and variance σ^2/n.

37. Let X_1 and X_2 be independent chi-squared random variables with 5 and 10 degrees of freedom respectively. Show that $X_1 + X_2$ is a chi-squared random variable with 15 degrees of freedom.

38. Let $X_1, X_2, X_3, \ldots, X_n$ be independent chi-squared random variables with $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n$ degrees of freedom respectively. Let

$$Y = X_1 + X_2 + X_3 + \cdots + X_n$$

Show that Y is a chi-squared random variable with γ degrees of freedom where $\gamma = \sum_{i=1}^{n} \gamma_i$.

39. It can be shown that the square of a standard normal random variable has a chi-squared distribution with $\gamma = 1$. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Use Exercise 38 to show that

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$

has a chi-squared distribution with n degrees of freedom.

40. Let X denote the time required to do a computation using an algorithm written in Fortran and let Y denote the time required to do the same calculation using an algorithm written in Pascal. Assume that X is normally distributed with mean 10 seconds and standard deviation 3 seconds that Y is normally distributed with mean 9 seconds and standard deviation 4 seconds.

(*a*) What is the distribution of the random variable $X - Y$?

(*b*) Find the probability that a given calculation will run faster using Fortran than when using Pascal.

Section 7.4

41. As heat is added to a material its temperature rises. The heat capacity is a quantitative statement of the increase in temperature for a specified addition of heat. These data are obtained on X, the measured heat capacity of liquid ethylene glycol at constant pressure and 80°C. Measurements are in calories per gram degree Celsius.

.645	.654	.640	.627	.626
.649	.629	.631	.643	.633
.646	.630	.634	.631	.651
.659	.638	.645	.655	.624
.658	.658	.658	.647	.665

Past experience indicates that $\sigma = .01$.

(*a*) Evaluate $\overline{X}$ for these data thereby obtaining an unbiased point estimate for μ.

(*b*) Assume that X is normally distributed. Find a 95% confidence interval for μ.

(*c*) Would you expect a 90% confidence interval for μ based on these data to be longer or shorter than the interval of part (*b*)? Explain. Verify your answer by finding a 90% confidence interval on μ. *Hint:* Begin by sketching a curve similar to that shown in Fig. 7.3 with $1 - \alpha = .90$ and $\alpha/2 = .05$.

(*d*) Would you expect a 99% confidence interval for μ based on these data to be longer or shorter than the interval of part (*b*)? Explain. Verify your answer by finding a 99% confidence interval for μ.

42. The late manifestation of an injury following exposure to a sufficient dose of radiation is common. These data are obtained on the variable X, the time in days that elapses between the exposure to radiation and the appearance of peak erythema (skin redness).

16	12	14	16	13	9	15	7
20	19	11	14	9	13	11	3
8	21	16	16	12	16	14	20
7	14	18	14	18	13	11	16
18	16	11	13	14	16	15	15

(*a*) Even though the time at which the peak redness appears is recorded to the nearest day, time is actually a continuous random variable. Sketch a stem-and-leaf diagram for these data. Does the diagram lend support to the assumption that X is normally distributed?

(*b*) Evaluate $\overline{X}$ for these data.

(*c*) Assume that $\sigma = 4$ and find a 95% confidence interval on the mean time to the appearance of peak redness. Would you be surprised to hear a claim that $\mu = 17$ days? Explain, based on the confidence interval found in part (*b*).

43. When fission occurs, many of the nuclear fragments formed have too many neutrons for stability. Some of these neutrons are expelled almost instantaneously. These observations are obtained on X, the number of neutrons released during fission of plutonium-239:

3	2	2	2	2	3	3	3
3	3	3	3	4	3	2	3
3	2	3	3	3	3	3	1
3	3	3	3	3	3	3	3
3	3	2	3	3	3	3	3

(*a*) Is X normally distributed? Explain.

(*b*) Estimate the mean number of neutrons expelled during fission of plutonium-239.

(*c*) Assume that $\sigma = .5$. Find a 99% confidence interval on μ. What theorem justifies the procedure you used to construct this interval?

(*d*) The reported value of μ is 3.0. Do these data refute this value? Explain.

44. (*Central Limit Theorem.*) Consider an infinite population with 25% of the elements having the value 1, 25% the value 2, 25% the value 3, and 25% the value 4. If X is the value of a randomly selected item, then X is a discrete random variable whose possible values are $1, 2, 3, 4$.

(*a*) Find the population mean μ and population variance σ^2 for the random variable X.
(*b*) List all of the 16 possible distinguishable samples of size 2 and for each calculate the value of the sample mean. Represent the value of the sample mean $\overline{X}$ using a probability histogram (use one bar for each of the possible values for $\overline{X}$). Note that although this is a very small sample, the distribution of $\overline{X}$ does not look like the population distribution and has the general shape of the normal distribution.
(*c*) Calculate the mean and variance of the distribution of $\overline{X}$ and show that, as expected, they are equal to μ and σ^2/n, respectively.

REVIEW EXERCISES

45. Consider the random variable X with density given by

$$f(x) = (1 + \theta)x^{\theta} \qquad 0 < x < 1 \qquad \theta > -1$$

(*a*) Show that $\int_0^1 f(x)\,dx = 1$ regardless of the specific value chosen for θ.
(*b*) Find $E[X]$.
(*c*) Find the method of moments estimator for θ.
(*d*) Find the method of moments estimate for θ based on these data:

.5 .3 .1 .1 .2

(*e*) Find the maximum likelihood estimator for θ.
(*f*) Find the maximum likelihood estimate for θ based on the data of part (*d*). Does this value agree with the method of moments estimate?

46. Consider the random variable X with density given by

$$f(x) = 1/\theta \qquad 0 < x < \theta$$

(*a*) Find $E[X]$.
(*b*) Find the method of moments estimator for θ. Is this estimator unbiased for θ?
(*c*) Find the method of moments estimate for θ based on these data:

1 .5 1.4 2.0 .25

47. Studies have shown that the random variable X, the processing time required to do a multiplication on a new 3-D computer, is normally distributed with mean μ and standard deviation 2 microseconds. A random sample of 16 observations is to be taken.
(*a*) What is the distribution of $\overline{X}$?
(*b*) These data are obtained:

42.65	45.15	39.32	44.44
41.63	41.54	41.59	45.68
46.50	41.35	44.37	40.27
43.87	43.79	43.28	40.70

Based on these data, find an unbiased estimate for μ.
(*c*) Find a 95% confidence interval for μ. Would you be surprised to read that the average time required to process a multiplication on this system is 42.2 microseconds? Explain, based on the confidence interval.

48. Let X denote the unit price of an 8-in floppy diskette. These observations are obtained from a random sample of 10 suppliers:

\$3.83	3.54	3.44	3.89	3.65
3.70	3.59	3.37	4.04	3.93

(a) Find an unbiased estimate for the mean price of these diskettes.
(b) Find an unbiased estimate for the variance in the price of these diskettes.
(c) Find the sample standard deviation. Is this an unbiased estimate for σ?
(d) Assume that X is normally distributed. Find the maximum likelihood estimate for σ^2. Does this agree with your answer to (b)?

49. (*Central Limit Theorem.*) In an attempt to approximate the proportion p of improperly sealed packages produced on an assembly line, a random sample of 100 packages is selected and inspected. Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th package selected is improperly sealed} \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the distribution of X_i?
(b) Based on the Central Limit Theorem, what is the approximate distribution of $\bar{X}$?
(c) When the experiment is conducted, we observe five improperly sealed packages. Find a point estimate for the proportion of improperly sealed packages being produced on this assembly line.

50. (*Central Limit Theorem.*) In a study of the size of various computer systems, the random variable X, the number of files stored, is considered. Past experience indicates that $\sigma = 5$. These data are obtained:

7	8	4	5	9	9
4	12	8	1	8	7
3	13	2	1	17	7
12	5	6	2	1	13
14	10	2	4	9	11
3	5	12	6	10	7

(a) Find an unbiased estimate for μ, the mean number of files per system.
(b) Based on the Central Limit Theorem, what is the approximate distribution of $\bar{X}$?
(c) Find an approximate 98% confidence interval on μ.
(d) In describing the size of such systems, an executive states that the average number of files exceeds 10. Does this statement surprise you? Explain.

51. Let X denote the time expended by a terminal user in a computing session (time from log on to log off). Assume that X is normally distributed with $\mu_X = 15$ minutes and $\sigma_X = 4$ minutes. Let Y denote the time required to access the system. Assume that Y is normally distributed with mean 1.5 minutes and $\sigma_Y = .5$ minutes. Assume that X and Y are independent.
(a) Find $m_X(t)$ and $m_Y(t)$.
(b) The random variable $T = X + Y$ denotes the total time required by the user to run his job. Find the moment generating function for T.

(*c*) What is the distribution of T?
(*d*) Find the probability that the total time required exceeds 20 minutes.

52. Let $X_1, X_2, \ldots, X_{100}$ be a random sample of size 100 from a gamma distribution with $\alpha = 5$ and $\beta = 3$.
(*a*) Find the moment generating function for $Y = \sum_{i=1}^{100} X_i$.
(b) What is the distribution of Y?
(*c*) Find the moment generating function for $\overline{X} = Y/n$.
(*d*) What is the distribution of $\overline{X}$?
(*e*) Use the Central Limit Theorem to approximate the probability that $\overline{X}$ is at most 14.

53. Consider the random variable X with density given by

$$f(x) = (1/\theta^2)xe^{-x/\theta} \qquad x > 0 \qquad \theta > 0$$

(*a*) What is the distribution of X?
(*b*) What is $E[X]$?
(*c*) Find the method of moments estimator for θ.
(*d*) Find the maximum likelihood estimator for θ based on a random sample of size n. Does this estimator differ from that found in part (*c*)?
(*e*) Estimate θ based on these data:

3	5	2	3	4
1	4	3	3	3

(*f*) Are the estimators found in parts (*c*) and (*d*) unbiased estimators for θ?

54. Let X be normally distributed with mean 2 and variance 25.
(*a*) What is the distribution of the random variable $(X - 2)/5$?
(*b*) What is the distribution of the random variable $[(X - 2)/5]^2$?
(*c*) Let $X_1, X_2, X_3, \ldots, X_{10}$ represent a random sample from the distribution of X. What is the distribution of the random variable

$$\sum_{i=1}^{10} \left(\frac{X_i - 2}{5} \right)^2$$

***55.** (*Central Limit Theorem.*) In this problem you will use the Central Limit Theorem to justify the normal approximation to the Poisson distribution given earlier. That is, you will show that a Poisson random variable X with parameter λs can be approximated using a normal random variable with mean and variance λs. To do so, let $Y_1, Y_2, Y_3, \ldots, Y_n$ be a random sample of size n from a Poisson distribution with parameter $\lambda s/n$.
(*a*) Use moment generating function techniques to show that

$$X = \sum_{i=1}^{n} Yi$$

has a Poisson distribution with parameter λs.
(*b*) Use the Central Limit Theorem to find the approximate distribution of $\overline{Y}$.
(*c*) Note that $n\overline{Y} = X$. Use this observation to argue that X is approximately normally distributed with mean λs and variance λs.

56. (*Central Limit Theorem.*) Consider the experiment of tossing a fair die once. Let X denote the number that occurs. Theoretically, X follows a discrete uniform distribution.

(*a*) Find the theoretical density, mean, and variance for X.

(*b*) Now consider an experiment in which the die is tossed 20 times and the results averaged. By the Central Limit Theorem, what is the theoretical mean and variance for the random variable $\bar{X}$?

(*c*) Perform the experiment of part (*b*) 25 times and record the value of $\bar{X}$ each time. (You will toss the die 500 times and obtain a data set that consists of 25 averages.) What shape should the stem-and-leaf diagram for these data assume? Explain. Construct a stem-and-leaf diagram for your data. Did the diagram take the shape that you expected?

(*d*) Approximately what value would you expect to obtain if you averaged the data of part (*c*)? Average your 25 observations on $\bar{X}$. Did the result come out as expected?

(*e*) Approximately what value would you expect to obtain if you found the sample variance for the data of part (*c*)? Explain. Find s^2 for your 25 observations on $\bar{X}$. Did the result come out as expected?

(*f*) If you were to construct 95% confidence intervals on μ based on each of the values of $\bar{X}$ found in part (*c*), approximately how many of them would you expect to contain the true value of μ? From your data, can you find an example of a confidence interval that does contain μ? of a confidence interval that does not contain μ?

CHAPTER 8

INFERENCES ON THE MEAN AND VARIANCE OF A DISTRIBUTION

We have seen how to estimate both the mean and variance of a distribution via point estimation. We have also seen how to generate a confidence interval for the mean of a normal distribution when its variance is assumed to be *known*. Unfortunately, in most statistical studies, the assumption that σ^2 is known is unrealistic. If it is necessary to estimate the mean of a distribution, then its variance is usually unknown also. In this chapter we turn our attention to the problem of making inferences on the mean and variance of a distribution when both of these parameters are assumed to be unknown. We begin by considering the construction of a confidence interval for σ^2.

8.1 INTERVAL ESTIMATION OF VARIABILITY

In Theorem 7.1.3, we showed that the statistic S^2 is an unbiased estimator for σ^2. To obtain a $100(1 - \alpha)\%$ confidence interval for σ^2, we need a random variable whose expression involves σ^2 and whose probability distribution is known. In Exercise 39, Chap. 7, we showed that the random variable $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2$ has a chi-squared distribution with n degrees of freedom. The next theorem shows that if the population mean μ is replaced by the sample mean $\overline{X}$, the resulting random variable $\sum_{i=1}^{n}(X_i - \overline{X})^2/\sigma^2$ follows a chi-squared distribution with $n - 1$ degrees of freedom. This theorem provides the random variable needed to construct a confidence interval for σ^2.

Theorem 8.1.1 (Distribution of $(n - 1)S^2/\sigma^2$). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. The random variable

$$(n-1)S^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/\sigma^2$$

has a chi-squared distribution with $n - 1$ degrees of freedom.

Proof. This argument does not constitute a rigorous proof of our theorem. However, it does suggest that the distribution of the random variable $(n - 1)S^2/\sigma^2$ is as stated. We begin by rewriting the random variable as a difference between two chi-squared random variables.

$$\begin{aligned}
(n-1)S^2/\sigma^2 &= \sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{\sigma^2} = \sum_{i=1}^{n}\frac{[(X_i-\mu) - (\overline{X}-\mu)]^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} - 2(\overline{X}-\mu)\sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma^2} + \frac{n(\overline{X}-\mu)^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} - 2(\overline{X}-\mu)\frac{\left(\sum_{i=1}^{n}X_i - n\mu\right)}{\sigma^2} + \frac{n(\overline{X}-\mu)^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} - \frac{2n(\overline{X}-\mu)^2}{\sigma^2} + \frac{n(\overline{X}-\mu)^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} - \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2
\end{aligned}$$

We now see that

$$(n-1)S^2/\sigma^2 + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$$

Note that the random variable $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ is standard normal. By Exercise 39 of Chap. 7 $[(\overline{X} - \mu)/(\sigma/\sqrt{n})]^2$ has a chi-squared distribution with one degree of freedom, and $\sum_{i=1}^{n}[(X_i - \mu)^2/\sigma^2]$ has a chi-squared distribution with n degrees of freedom. Since the sum of independent chi-squared random variables is also a chi-squared random variable (see Exercise 38 of Chap. 7), it is logical to assume that the random variable $(n - 1)S^2/\sigma^2$ has a chi-squared distribution with $n - 1$ degrees of freedom as claimed.

To use the random variable $(n - 1)S^2/\sigma^2$ to derive a $100(1 - \alpha)\%$ confidence interval on σ^2, we first partition the X^2_{n-1} curve as shown in Fig. 8.1. It is

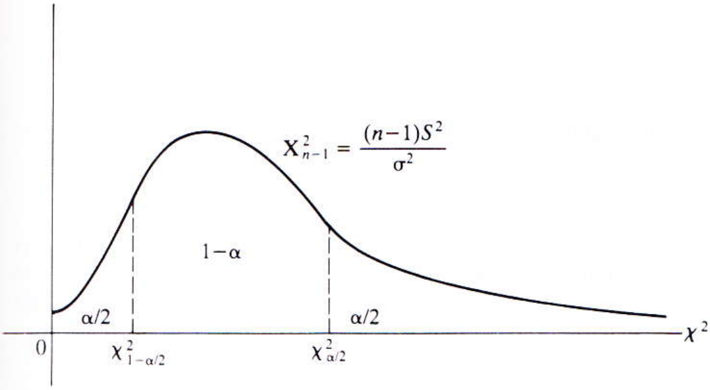


FIGURE 8.1
Partitions of the X^2_{n-1} curve needed to derive a $100(1-\alpha)\%$ confidence interval on σ^2.

evident that

$$P\left[\chi^2_{1-\alpha/2} \le (n-1)S^2/\sigma^2 \le \chi^2_{\alpha/2}\right] = 1 - \alpha$$

To find the lower and upper bounds for the confidence interval, we isolate σ^2 in the center of the inequality by inverting each term and solving for σ^2.

$$P\left[1/\chi^2_{\alpha/2} \le \sigma^2/(n-1)S^2 \le 1/\chi^2_{1-\alpha/2}\right] = 1 - \alpha$$

or

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha$$

The desired confidence bounds can be read from the latter inequality and are given in Theorem 8.1.2.

Theorem 8.1.2 ($100(1-\alpha)\%$ confidence interval on σ^2). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. The lower and upper bounds, L_1 and L_2 respectively, for a $100(1-\alpha)\%$ confidence interval on σ^2, are given by

$$L_1 = (n-1)S^2/\chi^2_{\alpha/2} \qquad \text{and} \qquad L_2 = (n-1)S^2/\chi^2_{1-\alpha/2}$$

As one would suspect, to obtain the bounds for a $100(1-\alpha)\%$ confidence interval on the standard deviation of a normal random variable, we take the nonnegative square root of the bounds given in Theorem 8.1.2.

Example 8.1.1. In computing, "work load" is defined as a collection of processor and input-output (I/O) resource requests during a particular period to time. Workloads are compared via a measure called relative I/O content. The average commercial batch MVS installation provides the base for this measure and is given a relative I/O content rating of 1. Other installations are rated relative to this base.

0	47
1	457486
2	0505
3	405611090
4	201
5	1

FIGURE 8.2
Stem-and-leaf diagram of the relative I/O content of the consulting firm of Example 8.1.1.

These observations on the relative I/O content for a large consulting firm over randomly selected one-hour periods are obtained: (Based on "Processor, I/O Path and DASD Configuration Capacity" by J. B. Major, *IBM Systems Journal*, vol. 20, no. 1, 1981, pp. 63–85.)

3.4	3.6	4.0	0.4	2.0
3.0	3.1	4.1	1.4	2.5
1.4	2.0	3.1	1.8	1.6
3.5	2.5	1.7	5.1	.7
4.2	1.5	3.0	3.9	3.0

Let us construct a 95% confidence interval on the standard deviation of the relative I/O content for this installation. The stem-and-leaf diagram for these data is shown in Fig. 8.2. This diagram does not suggest a serious departure from normality. The partition of the X^2_{24} curve needed to construct the confidence interval is shown in Fig. 8.3. For these data,

$$\Sigma x = 66.5 \qquad \Sigma x^2 = 210.67$$

$$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = 1.407 \qquad s = \sqrt{1.407} = 1.186$$

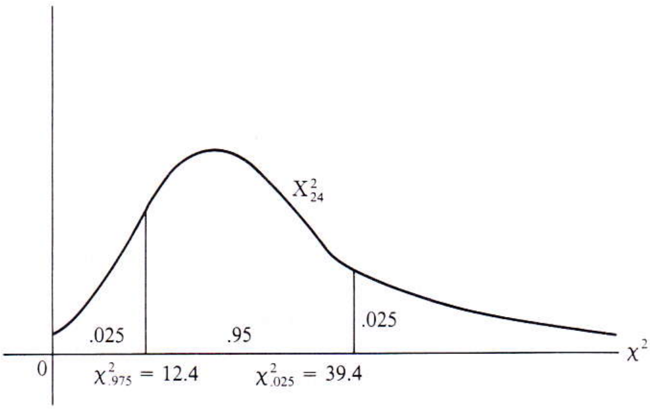


FIGURE 8.3
Partition of the X^2_{24} curve needed to construct a 95% confidence interval on the variance in relative I/O content of the consulting firm of Example 8.1.1.

The bounds for a 95% confidence interval on σ^2 are

$$L_1 = (n-1)s^2/\chi^2_{.025} = 24(1.407)/39.4 = .857$$

$$L_2 = (n-1)s^2/\chi^2_{.975} = 24(1.407)/12.4 = 2.723$$

The bounds for a 95% confidence interval on σ are

$$L_1 = \sqrt{.857} \doteq .926$$

$$L_2 = \sqrt{2.723} \doteq 1.650$$

8.2 ESTIMATING THE MEAN AND THE STUDENT-*t* DISTRIBUTION

Note that to obtain a point estimate for a population mean μ, it is not necessary to know the population variance; the sample mean $\overline{X}$ provides an unbiased estimator for μ regardless of the value of σ^2. However, the bounds for a $100(1-\alpha)\%$ confidence interval on μ given in Sec. 7.4 are $\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$. It is assumed that, even though the population mean is unknown, the population variance is known. Practically speaking, this assumption is not very realistic. In most instances, when a statistical study is being conducted, it is being done for the first time; there is no way to know prior to the study either the mean or the variance of the population of interest. We consider in this section the more realistic problem of constructing a confidence interval on a population mean when the population variance is assumed to be *unknown*.

To derive a general formula for a $100(1-\alpha)\%$ confidence interval on μ under these circumstances, it is natural to begin by considering the random variable used earlier, namely,

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

There are two problems to overcome:

1. The value of σ is not known and must be estimated.
2. The distribution of the random variable obtained by replacing σ by an estimator is not known.

The first problem is easy to overcome. We shall use the sample standard deviation S as an estimator for σ. The second problem is a little more difficult to solve. When we replace σ by its estimator S, the random variable $(\overline{X}-\mu)/(S/\sqrt{n})$ results. It can be shown that the distribution of this random variable is no longer standard normal. Rather, when sampling from a normal distribution, it follows what is called a Student-*t*, or simply a T distribution. This distribution was first described by W. S. Gosset in 1908. He used the pen name "Student" because his employers, an Irish brewery, did not want their competitors to know that they were using statistical methods in their work. We pause briefly to consider this distribution.

Definition 8.2.1 (T Distribution). Let Z be a standard normal random variable and let X_γ^2 be an independent chi-squared random variable with γ degrees of freedom. The random variable

$$T = \frac{Z}{\sqrt{X_\gamma^2/\gamma}}$$

is said to follow a T distribution with γ degrees of freedom.

This definition implies that to show that a random variable follows a T distribution we must show that it can be written as a ratio of a standard normal random variable to the square root of an independent chi-squared random variable divided by its degrees of freedom.

We note here the characteristics of T distributions that will be useful in the work that follows.

1. There are infinitely many T distributions, each identified by one parameter γ, called degrees of freedom. This parameter is always a positive integer. The notation T_γ denotes a T random variable with γ degrees of freedom.
2. Each T random variable is continuous. The density for a T random variable with γ degrees of freedom is given by
$$f(t) = \frac{\Gamma(\gamma+1)/2}{\Gamma(\gamma/2)\sqrt{\pi\gamma}}\left(1+\frac{t^2}{\gamma}\right)^{-(\gamma+1)/2} \qquad -\infty < t < \infty$$
3. The graph of the density of a T_γ random variable is a symmetric bell-shaped curve centered at 0.
4. The parameter γ is a shape parameter in the sense that as its value increases, the variance of the random variable T_γ decreases. Thus, as the value of γ increases, the bell curve associated with T_γ becomes more compact.
5. As the number of degrees of freedom increases, the bell curve associated with the T_γ random variable approaches the standard normal curve.

These ideas are illustrated in Fig. 8.4.

A partial summary of the cumulative distribution for selected values of γ is given in Table VI of App. A. The table is read just as the chi-squared table is read. That is, the degrees of freedom are listed as row headings, pertinent probabilities are listed as column headings, and the points associated with those probabilities are listed in the body of the table. We use our previous convention of denoting by t_r the point associated with the T_γ curve such that the area to the right of the point is r.

Example 8.2.1. Consider the random variable T_{10}.

1. From Table VI of App. A, $P[T_{10} \le 1.372] = F(1.372) = .90$. By our notational convention, $t_{.10} = 1.372$. [See Fig. 8.5(a).]

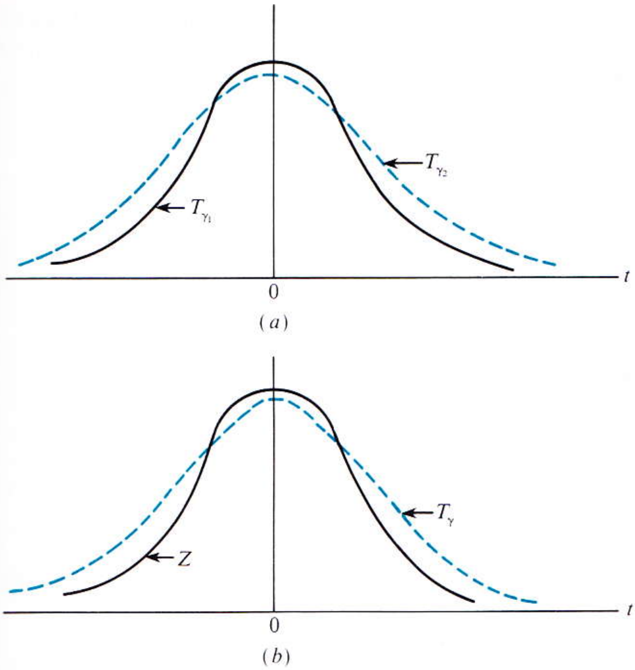
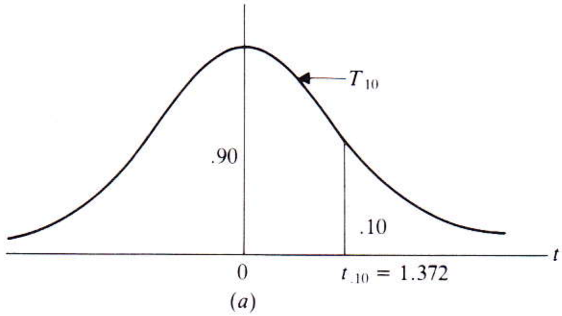


FIGURE 8.4
(*a*) Typical relationship between two T curves with $\gamma_1 > \gamma_2$. (*b*) Typical relationship between a T curve and the standard normal curve.

2. Due to the symmetry of the T curve, $t_{.90} = -t_{.10} = -1.372$.
3. The point t such that $P[-t \leq T_{10} \leq t] = .95$ is $t_{.025} = 2.228$. [See Fig. 8.5(*b*).]

The last row in Table VI of App. A is labeled ∞. The points listed in that row are actually points associated with the standard normal curve. Note that as γ increases, the values in each column of the table approach the value listed in the last row.

Let us now show that the random variable $(\bar{X} - \mu)/(S/\sqrt{n})$ follows a T distribution as claimed. The proof of this theorem depends on a result that is beyond the scope of this discussion mathematically. In particular, it can be shown that when sampling from a normal distribution, the sample mean $\bar{X}$ and the sample standard deviation S are independent. This result is not surprising. It says simply that knowledge of the center of location of a normal random variable does not contribute to knowledge of its variability. The next theorem provides the basis for the construction of a $100(1 - \alpha)\%$ confidence interval on μ when σ^2 is assumed to be unknown



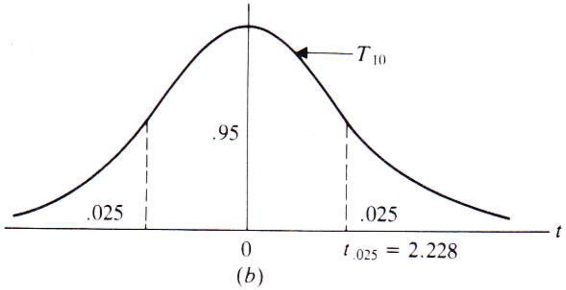


FIGURE 8.5
(a) $P[T_{10} \leq 1.372] = .90$. (b) $P[-2.228 \leq T_{10} \leq 2.228] = .95$.

Theorem 8.2.1. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. The random variable

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows a T distribution with $n - 1$ degrees of freedom.

Proof. We shall show that the random variable $(\bar{X} - \mu)/(S/\sqrt{n})$ can be written as the ratio of a standard normal random variable to the square root of an independent chi-squared random variable divided by its degrees of freedom. By Theorem 7.3.4, $\bar{X}$ is normal with mean μ and variance σ^2/n. Standardizing, $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is standard normal. By Theorem 8.1.1, $(n - 1)S^2/\sigma^2$ is a chi-squared random variable with $n - 1$ degrees of freedom. Consider the random variable

$$\frac{Z}{\sqrt{X_\gamma^2/\gamma}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)S^2/\sigma^2(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Since $\bar{X}$ and S are independent, this random variable follows a T distribution with $n - 1$ degrees of freedom as claimed.

It is now easy to determine the general form for a $100(1 - \alpha)\%$ confidence interval on μ when σ^2 is unknown. We need only note that the two random variables

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad \text{and} \qquad T_\gamma = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

have the same algebraic structure. Thus the algebraic argument given in Sec. 7.4 will hold with σ being replaced by S and $z_{\alpha/2}$ being replaced by $t_{\alpha/2}$. These substitutions result in Theorem 8.2.2.

Theorem 8.2.2 (100(1 − α)% Confidence interval on μ when σ^2 is unknown). Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. A 100(1 − α)% confidence interval on μ is given by

$$\bar{X} \pm t_{\alpha/2} S/\sqrt{n}$$

Example 8.2.2 illustrates the use of this theorem.

Example 8.2.2. Sulfur dioxide and nitrogen oxide are both products of fossil fuel consumption. These compounds can be carried long distances and converted to acid before being deposited in the form of "acid rain." These data are obtained on the sulfur dioxide concentration (in micrograms per cubic meter) in a Bavarian forest thought to have been damaged by acid rain. (Based on "Is Acid Deposition Killing West German Forests?" by Leslie Roberts, *Bioscience*, vol. 33, no. 5, May 1983, pp. 302–305.)

52.7	43.9	41.7	71.5	47.6	55.1
62.2	56.5	33.4	61.8	54.3	50.0
45.3	63.4	53.9	65.5	66.6	70.0
52.4	38.6	46.1	44.4	60.7	56.4

For these data,

$$\Sigma x = 1294 \qquad \bar{x} \doteq 53.92\ \mu g/m^3 \qquad s \doteq 10.07\ \mu g/m^3$$

$$\Sigma x^2 = 72102.2 \qquad s^2 \doteq 101.40$$

The partition of the T_{23} curve needed to find a 95% confidence interval on the mean sulfur dioxide concentration in this forest is shown in Fig. 8.6. The confidence bounds for the interval are

$$\bar{x} \pm t_{\alpha/2} s/\sqrt{n} = 53.92 \pm 2.069(10.07)/\sqrt{24}$$

That is, we are 95% confident that the mean sulfur dioxide concentration in this forest lies in the interval [49.67, 58.17]. The average concentration of this compound in undamaged areas of the country is 20 $\mu g/m^3$. Since this value is not included in

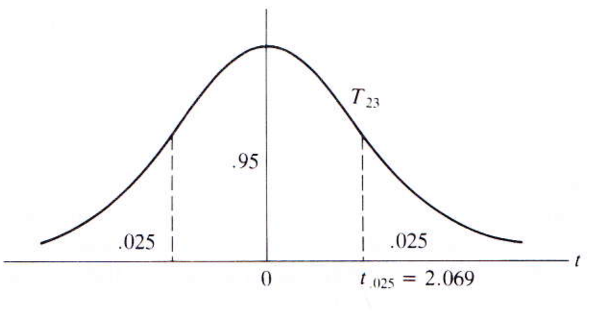


FIGURE 8.6
Partition of the T_{23} curve needed to construct a 95% confidence interval on the mean sulfur dioxide concentration in a Bavarian forest.

the above interval, there is evidence of an elevated sulfur dioxide concentration in the damaged forest.

Several things should be pointed out. First, the number of degrees of freedom involved in finding a confidence interval on μ when σ^2 is unknown is $n - 1$, the sample size minus 1. For large samples, this value may not be listed in Table VI of App. A. In this case, the last row in the table (∞) is used to find points of interest. Second, once again a normality assumption has been made. We can check the validity of this assumption graphically using a stem-and-leaf diagram or a histogram. More precise methods for testing for normality are available. One procedure that can be used is given in Sec. 8.6. If there is reason to suspect that the variable under study has a distribution that is not normal and the sample size is small then methods based on the T distribution may not be appropriate. Rather some nonparametric technique should be employed. Some of these techniques are discussed in Sec. 8.7.

8.3 HYPOTHESIS TESTING

We have considered the basic ideas of estimation in some detail. Recall that in a typical estimation problem there is some population parameter, θ, whose value is to be approximated based on a sample. Usually, there is *no* preconceived notion concerning the actual value of this parameter. We are attempting simply to ascertain its value to the best of our ability. In contrast, when testing a hypothesis on θ, there is a preconceived notion concerning its value. This implies that there are, in fact, two theories or hypotheses, involved in any statistical study of this sort: the hypothesis being proposed by the experimenter and the negation of this hypothesis. The former, denoted by H_1, is called the *alternative* or *research hypothesis*; the latter is denoted by H_0 and is called the *null hypothesis*. The purpose of the experiment is to decide whether the evidence tends to refute the null hypothesis. These three guidelines help in deciding how to state H_0 and H_1:

1. When testing a hypothesis concerning the value of some parameter θ, the statement of equality will always be included in H_0. In this way, H_0 pinpoints a specific numerical value that could be the actual value of θ. This value is called the "null value" and is denoted by θ_0.
2. Whatever is to be detected or supported is the alternative hypothesis.
3. Since our research hypothesis is H_1, it is hoped that the evidence leads us to reject H_0 and thereby accept H_1.

An example will help to clarify these ideas.

Example 8.3.1. Highway engineers have found that many factors affect the performance of reflective highway signs. One is the proper allignment of the automobile's headlights. It is thought that more than 50% of the automobiles on the road have misaimed headlights. If this contention can be supported statistically, then a new tougher inspection program will be put into operation. Let p denote the proportion of automobiles in operation that have misaimed headlights. Since we wish to support the statement that $p > .5$, this contention is taken as the alternative or research hypothesis, H_1. The null hypothesis is automatically the negation of H_1, namely, that $p \leq .5$. Thus the two hypotheses are

$$H_0: p \leq .5$$

$$H_1: p > .5$$

Note that the statement of equality appears in the null hypothesis. This pinpoints the value .5 as a possible value for p; that is, the "null value" for p is $p_0 = .5$. Note also that if H_0 is rejected, then our research hypothesis is accepted and the new inspection program will be implemented.

Once a sample has been selected and the data have been collected, a decision must be made. The decision will be either to reject H_0 or to fail to do so. The decision is made by observing the value of some statistic whose probability distribution is known *under the assumption that the null value is the true value of* θ. Such a statistic is called a *test statistic*. If the test statistic assumes a value that is rarely seen when $\theta = \theta_0$ and tends to lend credence to the alternative hypothesis, then we reject H_0 in favor of H_1; if the value observed is a commonly occurring one under the assumption that $\theta = \theta_0$, then we do not reject the null hypothesis. This means that at the end of any study we will be forced into exactly one of the following situations:

1. We will have rejected H_0 when it was true and will have committed what is known as a *Type I error*.
2. We will have made the correct decision of rejecting H_0 when the alternative H_1 was true.
3. We will have failed to reject H_0 when the alternative H_1 was true. In this case we will have committed what is known as a *Type II error*.
4. We will have made the correct decision of failing to reject H_0 when H_0 was true.

Example 8.3.2. In Example 8.3.1, we are testing

$$H_0: p \leq .5$$

$$H_1: p > .5 \quad \text{(majority of automobiles in operation have misaimed headlights)}$$

If a Type I error is made, we will have rejected H_0 when H_0 is true. Practically speaking, we will have concluded that a majority of cars on the road have misaimed headlights when, in fact, this is not true. This error could lead to the implementation of an unnecessary inspection program. A Type II error occurs if we fail to reject H_0 when H_1 is true. In this case, the inspection program would not be implemented when, in fact, it is needed.

Note that regardless of what is done, an error is possible. Anytime H_0 is rejected, a Type I error might occur; anytime H_0 is not rejected, a Type II error might occur. There is no way to avoid this dilemma. The job of the statistician is to design methods for deciding whether or not to reject H_0 that keep the probabilities of making either error reasonably small.

Philosophically there are two ways to determine whether or not to reject H_0. The first method, which we discuss in this section, is called *hypothesis testing*. This method has been used extensively in the past and is still used today. The second method, called *significance testing*, is becoming increasingly popular. It is discussed in the next section.

Hypothesis testing involves a procedure in which the values of the test statistic that lead to rejection of the null hypothesis are set before the experiment is conducted. These values constitute what is called the *critical, or rejection, region* for the test. The probability that the observed value of the test statistic will fall into this region by chance even though $\theta = \theta_0$ is called *alpha* (α), *the size of the test* or the *level of significance of the test*. If this occurs, a Type I error is committed. That is, in a hypothesis testing study, α is the probability of committing a Type I error. The next example illustrates the use of this method.

Example 8.3.3. To test the hypothesis of Example 8.3.1

H_0: $p \leq .5$

H_1: $p > .5$ (majority of automobiles in operation have misaimed headlights)

a random sample of 20 cars is selected and the headlights tested. Let us design a test so that α, the probability of rejecting H_0 when p is equal to the null value of .5 is about .05. The test statistic that we shall use is X, the number of cars in the sample with misaimed headlights. If p is, in fact, equal to the null value, then X is binomial with $n = 20$, $p = .5$, and $E[X] = np = 10$. Thus, if $p = .5$ then on the average, 10 of every 20 cars tested will have misaimed headlights; if H_1 is true this average value will be higher than 10. Logically, we should reject H_0 if the observed value of the test statistic X is somewhat larger than 10. Note from Table I of App. A that

$$\begin{aligned} P[X \geq 14|p = .5] &= 1 - P[X < 14|p = .5] \\ &= 1 - P[X \leq 13|p = .5] \\ &= 1 - .9423 \\ &= .0577 \end{aligned}$$

Let us agree to reject H_0 in favor of H_1 if the observed value of the test statistic, X, is 14 or greater. In this way we have split the possible values of X into two sets: $C = \{14, 15, 16, 17, 18, 19, 20\}$ and $C' = \{0, 1, 2, \ldots, 13\}$. If the observed value of X lies in C, we reject H_0 and conclude that the majority of cars in operation have misaimed headlights. The set of values of the test statistic that leads to rejection of the null hypothesis C is the *critical, or rejection, region* for the test. We chose C so that the probability that the test statistic will fall into C by chance, even though $p = .5$, is .0577. That is, we designed the test so that the probability of committing a Type I error (α) is approximately .05 as desired.

There is one point to note. In the previous example we use the null value $p_0 = .5$ to determine the critical region for the test even though the null hypothesis allows for values of p that are less than .5. It is safe to do this since values of X that are too large to occur by chance when $p = .5$ are also certainly too large to occur by chance when $p < .5$. That is, any value of X that leads us to reject .5 as a reasonable value for P also leads us to reject any value less than .5. (See Exercise 27.)

It is possible that the observed value of the test statistic does not fall into the rejection region even though H_0 is not true and should be rejected. If this occurs, a Type II error will be committed. The probability of this occurring is called *beta* (β). Beta is a little harder to handle than alpha, which can be dictated by the experimenter. For a particular test, β depends on the alternative. That is, β can be found only if a particular value of the alternative is specified. To illustrate, let us find β for the test designed in Example 8.3.3.

Example 8.3.4. The critical region for the test of Example 8.3.3 is $C = \{14, 15, 16, 17, 18, 19, 20\}$. Suppose that, unknown to the researcher, the true proportion of cars with misaimed headlights is .7. What is the probability that our test, as designed, is unable to detect this situation? To answer this question, we calculate β, the probability that H_0 will not be rejected given that $p = .7$. By definition,

$$\begin{aligned}
\beta &= P[\text{Type II error}] \\
&= P[\text{fail to reject } H_0 | p = .7] \\
&= P[X \text{ is not in the critical region} | p = .7] \\
&= P[X \leq 13 | p = .7] = .3920 \qquad \text{(Table I, App. A)}
\end{aligned}$$

That is, for the test as designed there is not a very high probability that we will be able to distinguish between $p = .5$ and $p = .7$. Beta is a function of the alternative in that if p is changed from .7 to .8, then β will change also. In this case

$$\beta = P[X \leq 13 | p = .8] = .0867$$

Note that as the difference between the null value of .5 and the alternative value of p increases, β decreases.

Remember that the hypothesis testing procedure entails deciding on the level of significance (α) before the data are gathered and the test statistic evaluated. That is, it involves presetting α. There are several reasons for wanting to do this. It gives a clearcut way of making a decision. Once α is set, the critical region for the test is fixed also. If the observed value of the test statistic falls into this region, we reject H_0: otherwise, we do not. There is no room for debate after the data are gathered. Hence there can be no charge that the statisticians are manipulating the results to suit themselves. In addition, if the consequences of making a Type I error are very serious, then by presetting α we are able to specify *before the fact* exactly how large a risk we are willing to tolerate. The language underlying hypothesis testing is summarized in Fig. 8.7.

	Actual situation	
Decision	H_0 true	H_1 true
Reject H_0	Type I error (probability α)	Correct decision
Fail to reject H_0	Correct decision	Type II error (probability β)

FIGURE 8.7

8.4 SIGNIFICANCE TESTING

In the last section we considered a method for deciding whether or not to reject a null hypothesis called *hypothesis testing*. In this section we consider another method for doing so. This method, called *significance testing*, is coming into widespread use. This is due to its logical appeal and to the increasing use of computer packages in analyzing statistical data.

To understand why significance testing is so appealing, let us point out a bothersome aspect of hypothesis testing that might have occurred to you already. It is easy to spot the problem with a simple example. Suppose that we want to test

$$H_0\colon p \le .1$$
$$H_1\colon p > .1$$

based on a sample of size 20. The test statistic is X, the number of "successes" that are observed in the 20 trials. Since the null value is $p_0 = .1$, the test statistic follows a binomial distribution with $E[X] = np_0 = 20(.1) = 2$. Values of X somewhat larger than 2 tend to lend credence to the alternative hypothesis. Suppose that we want α to be "very small" so we define the critical region to be $C = \{9, 10, 11, \ldots, 20\}$. For this test

$$\begin{aligned}
\alpha &= P[\text{Type I error}] \\
&= P[\text{reject } H_0 | p = p_0] \\
&= P[X \text{ is in the critical region} | p = .1] \\
&= P[X \ge 9 | p = .1] \\
&= 1 - P[X < 9 | p = .1] \\
&= 1 - P[X \le 8 | p = .1] \\
&= 1 - .9999 \\
&= .0001
\end{aligned}$$

This is indeed a "very small value"! Now suppose that we conduct our test and observe 8 "successes." Via our rather rigid rules for hypothesis testing, we are unable to reject H_0 since 8 does not lie in the critical region. However, a little thought should make you a bit uneasy with this decision! Note that 8 is very close to 9, our rather arbitrarily selected lower boundary for the critical region.

Let us see what the chances are of obtaining a value of 8 or more when $p = .1$.

$$\begin{aligned} P[X \geq 8 | p = .1] &= 1 - P[X < 8 | p = .1] \\ &= 1 - P[X \leq 7 | p = .1] \\ &= 1 - .9996 \\ &= .0004 \end{aligned}$$

This probability is certainly also "very small." It is hard to imagine a situation in which we would be willing to tolerate 1 chance in 10,000 of making a Type I error but would declare vehemently that 4 chances in 10,000 of making such an error is much too large to risk! There is so little difference between these probabilities that it seems a bit silly to insist that we adhere rigidly to our original cutoff point of 9.

The problem just demonstrated can be avoided by performing what is called a *significance test* rather than a hypothesis test. This method of deciding whether or not to reject H_0 entails setting up H_0 and H_1 exactly as before. However, we do not then preset α and specify a rigid critical region. Rather, we evaluate the test statistic and then determine the probability of observing a value of the test statistic at least as extreme as the value noted under the assumption that $\theta = \theta_0$. This probability is referred to by a variety of names, including the *critical level*, the *descriptive level of significance*, and the *probability*, or *P value* of the test. We use the term "*P* value" in this text. Note that the *P* value is the smallest level at which we could have preset α and still have been able to reject H_0. We reject H_0 if we consider this *P* value to be small.

Example 8.4.1. Automotive engineers are using more and more aluminum in the construction of automobiles in hopes of reducing the cost and improving gas mileage. For a particular model, the number of miles per gallon obtained on the highway currently has a mean of 26 mpg with a standard deviation of 5 mpg. It is hoped that a new design, which utilizes more aluminum, will increase the mean mileage rating. Assume that σ is not affected by this change. Since our research hypothesis is taken as the alternative hypothesis, we are testing

H_0: $\mu \leq 26$

H_1: $\mu > 26$ (the new design increases gas mileage on the highway)

Since the sample mean is an unbiased estimator for the population mean, a logical test statistic is $\bar{X}$. Let us agree to reject H_0 in favor of H_1 if the observed value of the sample mean is "somewhat larger" than 26. By "somewhat larger" we mean too large to have reasonably occurred by chance if the true mean highway mileage is still 26 mpg. These data are obtained during road testing:

33.8	24.3	18.8	23.7	25.3	29.6
24.9	31.5	34.4	28.0	20.5	36.7
30.3	33.5	27.4	27.6	22.5	30.7
28.6	27.1	28.8	16.5	32.7	25.2
33.1	37.5	25.1	34.5	29.5	26.8
30.0	28.4	25.6	19.8	28.9	27.7

The sample mean for these data is $\bar{x} = 28.04$ mpg. This value is larger than the null value for μ of 26 mpg. To see if there is enough difference to cause us to reject H_0, we find the P value for the test. That is, we compute the probability of observing a sample mean of 28.04 or larger if $\mu = 26$ and $\sigma = 5$. This is done by noting if $\mu = 26$ and $\sigma = 5$ then the test statistic $\bar{X}$ is, by the Central Limit Theorem (Theorem 7.4.2), at least approximately normally distributed with mean $\mu = 26$ and standard deviation $\sigma/\sqrt{n} = 5/6$. Therefore

$$\begin{aligned}
P[\bar{X} \geq 28.04 | \mu = 26, \sigma = 5] &= P\left[\frac{\bar{X} - 26}{(5/6)} \geq \frac{28.04 - 26}{(5/6)}\right] \\
&\doteq P[Z \geq 2.45] \\
&= 1 - P[Z \leq 2.45] \\
&= 1 - .9929 \qquad (\text{Table V, App. A}) \\
&= .0071
\end{aligned}$$

There are two explanations for this very small probability. The null hypothesis is true and we have observed a very rare sample that *by chance* has a large sample mean; the null hypothesis is not true and the new process has in fact resulted in a higher mean mileage rating. We prefer the latter explanation! That is, we shall reject H_0 and report that the P value of our test is .0071.

Significance testing is an interesting and exciting concept. There are still some questions to be answered concerning its proper use. One question still to be resolved is: "How do we compute a P value for a two-tailed test?" If the distribution of the test statistic is symmetric, as it is for a Z or T statistic, then it is logical to double the apparent one-tailed P value. If the distribution is not symmetric, as with a chi-squared statistic, then presumably the two-tailed P value is nearly double the one-tailed value. This is only one of several proposed solutions to the problem but it is the convention that we shall use. The reader is referred to [16] for an excellent discussion of the problem.

8.5 HYPOTHESIS AND SIGNIFICANCE TESTS ON THE MEAN

One of the most commonly encountered problems is that of testing a hypothesis concerning the value of the mean. We have seen how this can be done if it is assumed that σ^2 is known. Since this assumption is usually not valid, we turn our attention to a method that can be used to test hypotheses concerning μ when σ^2 is unknown and must be estimated from the data at hand. Consider these examples.

Example 8.5.1. The maximum acceptable level for exposure to microwave radiation in the United States is an average of 10 microwatts per square centimeter. It is feared that a large television transmitter may be polluting the air nearby by pushing the level of microwave radiation above the safe limit. Since our research hypothesis is taken as the alternative, we are testing

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10 \qquad \text{(unsafe)}$$

Example 8.5.2. Design engineers are working on a low-effort steering system that can be used in vans modified to fit the needs of disabled drivers. The old-type steering system required a force of 54 ounces to turn the van's 15-inch-diameter steering wheel. It is hoped that the new design will reduce the average force required to turn the wheel. In this case we are testing

$$H_0: \mu \geq 54$$
$$H_1: \mu < 54 \qquad \text{(new system requires less force to operate than the old)}$$

(*Design News*, April 1983, pp. 14–16.)

Example 8.5.3. A computer system currently has 10 terminals and uses a single printer. The average turnaround time for the system is 15 minutes. Ten new terminals and a second printer are added to the system. We want to determine whether or not the mean turnaround time is affected. To decide, we want to test

$$H_0: \mu = 15$$
$$H_1: \mu \neq 15 \qquad \text{(the new equipment has an impact on turnaround time)}$$

As you can see, a hypothesis on μ can take one of three general forms. With μ_0 denoting the null value of the mean, these are:

I	II	III
H_0: $\mu \leq \mu_0$	H_0: $\mu \geq \mu_0$	H_0: $\mu = \mu_0$
H_1: $\mu > \mu_0$	H_1: $\mu < \mu_0$	H_1: $\mu \neq \mu_0$
Right-tailed test	Left-tailed test	Two-tailed test

Form I is called a right-tailed test because when a hypothesis of this form is tested, the natural critical region is the upper- (or right-)tail region of the distribution of the test statistic. This point is explained in Example 8.5.4. Similarly form II is a left-tailed test because the natural critical region is the lower- (or left-)tail region of the appropriate distribution. In a two-tailed test, the critical region consists of both the lower- and upper-tail regions of the distribution of the test statistic. This is easy to remember because in a one-sided test, forms I and II, the inequality in the *alternative* hypothesis points toward the critical region.

There is one general statement to keep in mind when you test a hypothesis on any parameter:

> To test a hypothesis on a parameter θ, you must find a statistic whose probability distribution is known at least approximately under the assumption that $\theta = \theta_0$.

This statistic will serve as a test statistic. In the case at hand, such a statistic is easy to find. From the discussion of Sec. 8.2, we know that if X is normal the statistic $(\bar{X} - \mu_0)/(S/\sqrt{n})$ follows a T_{n-1} distribution. Tests based on this statistic are commonly called "T tests."

Tests of hypotheses on μ are actually conducted by testing H_0: $\mu = \mu_0$ against one of the alternatives $\mu > \mu_0$, $\mu < \mu_0$ or $\mu \neq \mu_0$. It is safe to do this for reasons analogous to those discussed in Sec. 8.3. In particular, values of the test statistic that lead us to reject μ_0 and conclude that $\mu > \mu_0$ will also lead us to reject any value less than μ_0; values of the test statistic that lead us to reject μ_0 and conclude that $\mu < \mu_0$ will also lead us to reject any value greater than μ_0. For this reason many statisticians prefer to express the three forms as

I	II	III
H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
H_1: $\mu > \mu_0$	H_1: $\mu < \mu_0$	H_1: $\mu \neq \mu_0$
Right-tailed test	Left-tailed test	Two-tailed test

This emphasizes the fact that when performing a hypothesis test on μ, α is computed assuming that $\mu = \mu_0$; when performing a significance test on μ, the P value is computed under the assumption that $\mu = \mu_0$. We shall follow this notational convention in the remainder of this text.

Example 8.5.4. To determine whether a large television transmitter is polluting the nearby air (see Example 8.5.1), we intend to test

$$H_0: \mu = 10$$
$$H_1: \mu > 10$$

A sample of 25 readings is to be obtained at randomly selected times over a one-week period. Our test statistic, $(\bar{X} - 10)/(S/\sqrt{25})$, follows a T_{24} distribution if H_0 is true. Since $\bar{X}$ is an unbiased estimator for the mean, we expect the observed value of $\bar{X}$ to be close to 10 if H_0 is true. This forces the numerator of the test statistic, $(\bar{X} - 10)$, to be small, causing the observed value of the test statistic to be small also. However, if H_1 is true, we expect $\bar{X}$ to be larger than 10, forcing $\bar{X} - 10$ to be large and *positive*. This in turn, results in a *large positive* value for the test statistic. Hence logically we should reject H_0 in favor of H_1 whenever the observed value of the test statistic is positive and too large to have reasonably occurred by chance. Thus the natural critical region for the test is the right-tail, or upper, region of the T_{24} distribution. To decide how large a value is needed in order to reject H_0, let us preset α. If we make a Type I error, we will shut down the transmitter unnecessarily; if we make a Type II error, we will fail to detect a potential health hazard. We want α small but not so small as to force β to be extremely large. Let us choose α to be .1. The critical point for the test, read from Table VI of App. A and shown in Fig. 8.8, is 1.318. We shall reject H_0 in favor of H_1 if the observed value of the test statistic is 1.318 or larger. When the experiment is conducted, it is found that $\bar{x} = 10.3$ and $s = 2$. The observed value of the test statistic is

$$(\bar{x} - 10)/(s/\sqrt{25}) = (10.3 - 10)/(2/5) = .75$$

Since this value falls below the critical point of 1.318, we are unable to reject H_0. These data do not support the contention that the transmitter is forcing the average microwave level above the safe limit.

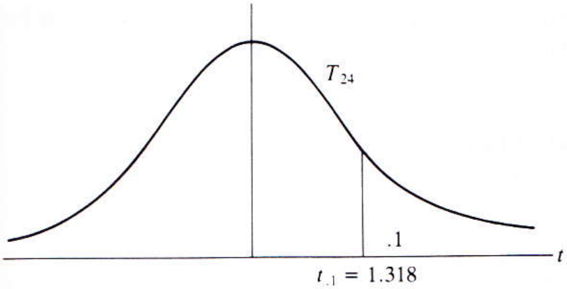


FIGURE 8.8
Critical region for an $\alpha = .1$ level right-tailed test ($n = 25$).

The next example illustrates the use of significance testing in testing a two-tailed hypothesis.

Example 8.5.5. In studying the effect of adding 10 new terminals and one printer to an existing computer system (see Example 8.5.3) we are testing

$H_0: \mu = 15$

$H_1: \mu \neq 15$ (the new equipment has an impact on turnaround time)

Since we do not have any *preconceived* notion as to whether the new equipment increases or decreases the mean turnaround time, we are conducting a two-tailed test. We shall reject H_0 in favor of H_1 if the observed value of the test statistic is too large in either the positive or negative sense to have occurred by chance. When the data are gathered, a sample of size 30 yields $\bar{x} = 14.0$ and $s = 3$. The observed value of the test statistic is

$$(\bar{x} - 15)/(s/\sqrt{30}) = (14 - 15)/(3/\sqrt{30}) \doteq -1.83$$

From Table VI of App. A, we see that

$$P[T_{29} \leq -1.699] = .05 \qquad \text{and} \qquad P[T_{29} \leq -2.045] = .025$$

Since -1.83 lies between -1.699 and -2.045, the probability of observing a value as large in the negative sense as that observed lies between .025 and .05. However, we were running a two-tailed test. This means that the P value of the test is the probability of observing a value as extreme as that observed in either the *positive or the negative* sense. That is, the P value is assumed to be double that computed above. We can report that for this test, $.05 < P < .1$. Since this probability is still small, we reject H_0 and conclude that the new equipment does affect the mean turnaround time.

It should be emphasized that the statistic $(\bar{X} - \mu_0)/(S\sqrt{n})$ follows the T_{n-1} distribution if X is *normal*. If X is not normal then care must be taken. It has been found that for samples of moderate to large size ($n \geq 25$), violating this assumption does not seriously affect the distribution of the test statistic in that the probability of committing a Type I and a Type II error is not appreciably changed [6]. This property is called "robustness." However, if the sample size is small then T tests should not be run on nonnormal data. Rather some nonparametric method should be used. Several of these are presented in Sec. 8.7. We present a quick method for testing for normality in Sec. 8.6 so that you can be sure that the statistic used to test a hypothesis is appropriate.

8.6 TESTING FOR NORMALITY AND HYPOTHESIS TESTS ON THE VARIANCE (OPTIONAL)

As has been pointed out, the methods discussed in the previous sections of this chapter assume that sampling is from a normal distribution. For many years after the discovery of the normal curve, practitioners felt that virtually every random variable was at least approximately normally distributed. As more and more data became available it became evident that this was not true. However, the statistical tools developed by Fisher, Pearson, and "Student," which presuppose normality, were so appealing to researchers in a wide variety of fields that they were eager to adopt them. To laymen, unable to follow the mathematical derivations of these techniques, the normality "assumption" was thought to be unimportant, a law of nature, or at best satisfied due to some sophisticated mathematical magic. The situation is best described by the words of Lippman in a remark to Poincare (1912) [6].

> Everyone believes in it (normal law of errors) however, said Monsieur Lippman to me one day, for the experimenters fancy that it is a theorem in mathematics and the mathematicians that it is an experimental fact.

Thus far, we have done only a rough check of the normality assumption using the stem-and-leaf diagram. If the stem-and-leaf diagram assumes a definite bell shape, then the normality assumption is probably reasonable. However, if there is a doubt in your mind concerning its validity, then you should test to see if there is statistical evidence that the data are drawn from a nonnormal distribution. Several methods are available for doing so. Here we consider a graphical method that is especially useful when sample sizes are small. A nongraphical procedure for use with large samples is in SAS supplement III. If you wish to skip the discussion here and go on to the subsection on hypothesis tests on the variance you can do so. The exercises for this section are flexible enough to allow this.

Testing for Normality

The method for detecting nonnormality that we consider here, called "the Lilliefors test for normality," was developed by H. W. Lilliefors in the late 1960s. Although it can be used with large samples, it is most helpful when samples are relatively small. The test basically compares the observed relative cumulative frequency distribution of the sample to that of the standard normal distribution. This is done by graphing the observed distribution on a Lilliefors graph. Figure

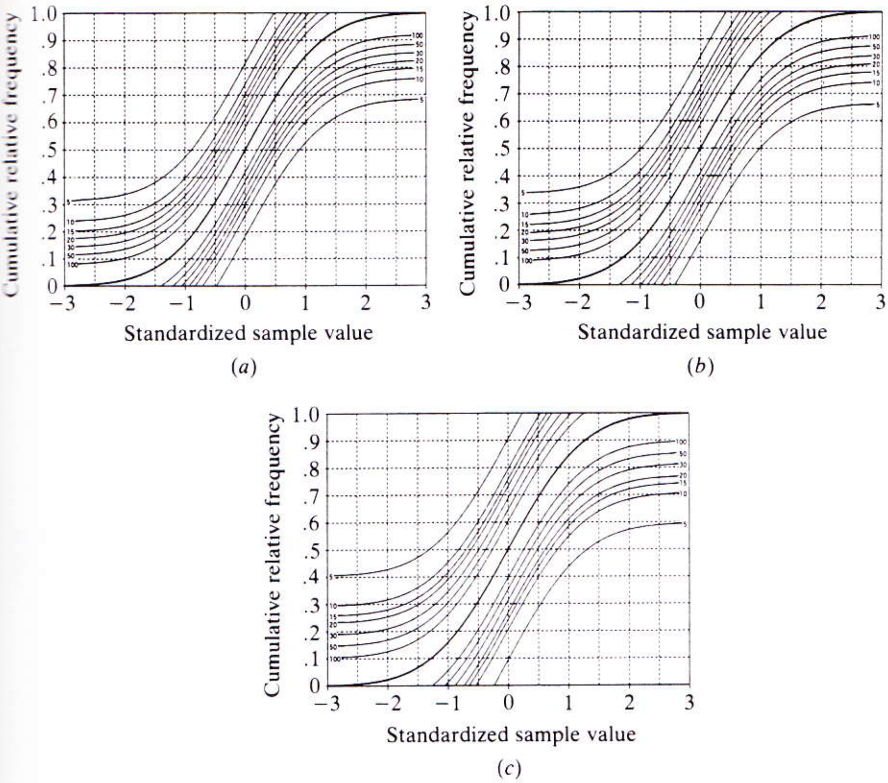


FIGURE 8.9
(*a*) 90% Lilliefors bounds for normal samples ($\alpha = .1$). (*b*) 95% Lilliefors bounds for normal samples ($\alpha = .05$). (*c*) 99% Lilliefors bounds for normal samples ($\alpha = .01$).

8.9 gives the Lilliefors graphs needed to test

$$H_0\text{: data are from a normal distribution}$$

$$H_1\text{: data are not from a normal distribution}$$

for various significance levels. The heavy curve in the center of the graph is the cumulative distribution for the standard normal curve. Curves to either side represent the Lilliefors bounds for the sample sizes indicated. If the observed relative cumulative frequency falls outside the bounds given for the specified sample size, then H_0 is rejected and it is concluded that the data are not from a normal distribution. The use of this technique is illustrated in Example 8.6.1.

Example 8.6.1. One random variable studied while designing the front-wheel-drive half-shaft of a new model automobile is the displacement (in millimeters) of the constant velocity (CV) joints. With the joint angle fixed at 12°, twenty simulations

TABLE 8.1

Observation	Standardized observation	Relative cumulative frequency
1.1	−1.62	.05
1.3	−1.48	.10
1.4	−1.41	.15
1.5	−1.34	.20
1.9	−1.06	.25
2.5	−.63	.30
2.6	−.56	.35
3.2	−.13	.40
3.5	.08	.45
3.7	.22	.50
3.7	.22	.50
3.9	.36	.60
4.1	.50	.65
4.2	.57	.70
4.2	.57	.70
4.4	.72	.80
4.6	.86	.85
4.8	1.00	.90
4.9	1.07	.95
6.2	1.99	1.00

were conducted. These data result: ("Closed Loop." *Magazine of Mechanical Testing*, June 1980, p. 18)

6.2	1.9	4.4	4.9	3.5
4.6	4.2	1.1	1.3	4.8
4.1	3.7	2.5	3.7	4.2
1.4	2.6	1.5	3.9	3.2

For these data, $\bar{x} = 3.39$ and $s = 1.41$. Since we are to compare the observed relative cumulative frequency distribution to that of the *standard* normal distribution, we first "standardize" these observations by subtracting $\bar{x}$ and dividing by s. The standardized observations are then ordered from smallest to largest and the relative cumulative frequency for each observation is found. The results of these calculations are shown in Table 8.1. To use these data to test

H_0: data are from a normal distribution

H_1: data are not from a normal distribution

at the $\alpha = .05$ level, we graph the observed relative cumulative frequency on the Lilliefors graph of Fig. 8.9(b). The result is given in Fig. 8.10. Since the graph of the

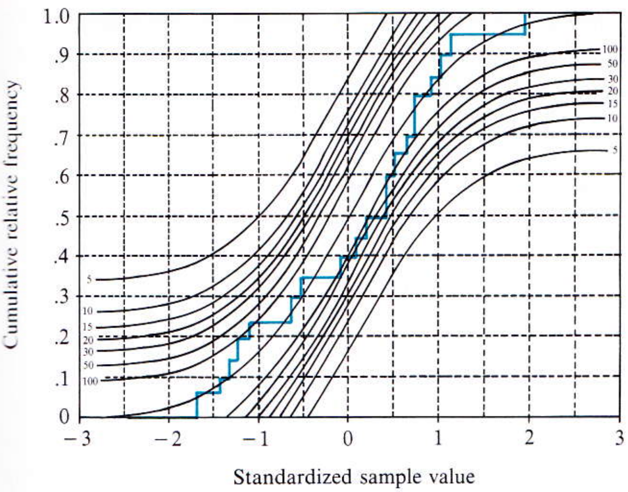


FIGURE 8.10
Test for normality of displacement of the CV joint at the $\alpha = .05$ level.

observed relative cumulative frequency does not fall outside the bands labeled 20, the size of our sample, we are unable to reject H_0. We have no evidence that the data are drawn from a nonnormal distribution.

Testing Hypotheses on the Variance

We now turn our attention to testing hypotheses on the value of σ^2 or σ. These tests take the same general form as tests on the mean. These are summarized below with σ_0^2 denoting the null value of the population variance.

I		II		III	
	H_0: $\sigma^2 = \sigma_0^2$		H_0: $\sigma^2 = \sigma_0^2$		H_0: $\sigma^2 = \sigma_0^2$
	H_1: $\sigma^2 > \sigma_0^2$		H_1: $\sigma^2 < \sigma_0^2$		H_1: $\sigma^2 \neq \sigma_0^2$
	Right-tailed test		Left-tailed test		Two-tailed test

The test statistic used to test each of these is $(n - 1)S^2/\sigma_0^2$. When sampling from a normal distribution, this statistic is known to follow a chi-squared distribution with $n - 1$ degrees of freedom provided $\sigma^2 = \sigma_0^2$. As expected, the critical regions for right- and left-tailed tests are the upper- and lower-tail regions of the X_{n-1}^2 distribution respectively; the critical region for the two-tailed test consists of both the upper- and lower-tail regions of the distribution.

Example 8.6.2. Engineers designing the front-wheel-drive half-shaft described in Example 8.6.1 claim that the standard deviation in the displacement of the CV shaft is less than 1.5 mm. The estimated standard deviation based on the given 20 observations is 1.41 mm. Do these data support the contention of the engineers? To answer this question, we test

$$H_0: \sigma = 1.5$$

$$H_1: \sigma < 1.5$$

This is equivalent to testing

$$H_0: \sigma^2 = (1.5)^2$$

$$H_1: \sigma^2 < (1.5)^2$$

The observed value of the test statistic is

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{19(1.41)^2}{(1.5)^2} = 16.79$$

Since the test is left-tailed, we reject H_0 if this value is too small to have occurred by chance when H_0 is true. From the chi-squared table, we see that

$$P[X_{19}^2 \le 14.6] = .25 \qquad \text{and} \qquad P[X_{19}^2 \le 18.3] = .50$$

Since the observed value of the test statistic, 16.79, lies between 14.6 and 18.3, the P value of the test lies between .25 and .50. Since this P value is rather large, we are unable to reject H_0. These data are not sufficient to allow us to claim that $\sigma < 1.5$ mm.

Recall that when sample sizes are moderate to large ($n \ge 25$), the T statistic can be used to make inferences on μ even though the normality assumption may be violated. It is when sample sizes are small that this becomes a serious problem. Unfortunately, the same cannot be said concerning the use of the X_{n-1}^2 statistic for making inferences on σ^2 and σ. For this reason, when constructing confidence intervals on σ^2 or testing hypotheses on the value of this parameter, a check for normality must be made. If the data are nonnormal then these methods should not be used.

8.7 ALTERNATIVE NONPARAMETRIC METHODS

We have seen how to use the Z and T statistics to test hypotheses concerning the mean of a normal distribution. The procedures presented assume either that we are sampling from a normal distribution or that sample sizes are large enough so that deviations from the normality assumption do not seriously affect our results. In reality, experimenters often obtain data for which it is clearly unreasonable to assume an underlying normal distribution and for which sample sizes are small. When this occurs, usually the experimenter is advised to use a "nonparametric" test for location rather than the usual Z or T test. In this section, we examine the meaning of the term "nonparametric" test. We also present some nonparametric alternatives for the usual Z and T tests for location.

The terms "nonparametric" and "distribution free" are often used interchangeably. When we use the term "nonparametric test," we shall mean a test with the property that no assumption is being made concerning the specific distribution from which the sample is drawn. Although we usually assume that the distribution is continuous, we do not have to specify the family to which the random variable under study belongs. In particular, we will no longer have to assume that the random variable being studied is normally distributed. Hence, nonparametric methods are applicable to a larger class of distributions than their normal theory analogs.

When comparing two statistical procedures designed to test essentially the same thing, we look at two characteristics: the probability of committing a Type I error and the power of the test. We want α to be small but at the same time we want a high probability of rejecting a false null hypothesis. Typically, for a fixed α level, the normal theory procedures are more powerful than their nonparametric counterparts *when the assumptions underlying the normal theory test are met*. However, studies have shown that when these assumptions are not met, the use of normal theory procedures leads to tests that are approximate in the sense that the apparent α level is suspect. For example, if we run a chi-squared test for variance on data that is far from normal at an apparent α level of .05, the actual probability of rejecting a true null hypothesis may be far from .05. In some cases the approximations are excellent, but in others they are so bad as to be completely unacceptable. In any case, using a normal theory procedure in situations in which the normal theory assumptions are not valid is dangerous. In such cases we turn to nonparametric procedures. These methods are usually superior for analyzing data when the normal theory assumptions are not met; they compare very favorably to the normal theory tests even when the normal theory assumptions are met. The safe course of action is to follow the advice: when in doubt use a nonparametric test!

In this section we will discuss the sign test and the Wilcoxon Signed-Rank test, both of which can be used to test for location in the form of population medians.

Sign Test for Median

Recall that for a continuous distribution the median for a random variable X is defined to be the value M such that

$$P(X < M) = P(X > M) = 1/2.$$

That is, the median is the 50th percentile of the distribution. For a symmetric distribution such as the normal, the population mean and median are identical. We will see that the sign test is simply a form of the binomial test which was discussed in Sec. 8.3. Let X denote a continuous random variable with median M and let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from this unspecified distribution. If M_0 denotes the hypothesized value of the population median, then the usual forms of the hypothesis to be tested can be stated as follows:

H_0: $M = M_0$	H_0: $M = M_0$	H_0: $M = M_0$
H_1: $M > M_0$	H_1: $M < M_0$	H_1: $M \neq M_0$
Right-tailed test	Left-tailed test	Two-tailed test

Under the assumption of a continuous distribution, the differences $X_i - M_0$ each have probability 1/2 of being positive, probability 1/2 of being negative, and probability 0 of being zero.

Let Q_+ denote the number of positive differences obtained. If H_0 is true, Q_+ is binomially distributed with parameters n and 1/2 and the expected value of Q_+ is $n/2$. That is, if H_0 is true half the differences should be positive and the rest negative. Note that in running a left-tailed test we want to detect a situation in which the true median M lies below the hypothesized median M_0. If this is true, we expect more than half of the differences to be negative. This creates fewer positive differences than expected. Thus a logical procedure is to reject H_0: $M = M_0$ in favor of H_1: $M < M_0$ if the observed value of Q_+ is too small to have occurred by chance. In conducting a right-tailed test the situation is reversed. In this case we reject H_0: $M = M_0$ in favor of H_1: $M > M_0$ if the observed value of Q_-, the number of negative differences obtained, is too small to have occurred by chance. A two-tailed test is conducted by rejecting H_0: $M = M_0$ in favor of H_1: $M \neq M_0$ if the smaller of Q_+ and Q_- is too small to have occurred by chance. The next example illustrates the use of the sign test.

Example 8.7.1. A standard method for completing a task on an assembly line yields a median completion time of 55 seconds. A new procedure is developed that should reduce the median time required. We want to test

$$H_0: M = 55$$
$$H_1: M < 55$$

To do so, 15 subjects are asked to complete the task and these observations are obtained on the random variable X, the time required.

35	65	48	40	70	50	58	36
47	41	49	39	34	33	31	

The stem-and-leaf diagram for these data is shown in Fig. 8.11. Note that the diagram does suggest that X is not normally distributed. Since the sample size is rather small, we will test for location using the nonparametric sign test. The test is left-tailed. Hence the test statistic is Q_+, the number of positive differences obtained

3	569431
4	80719
5	08
6	5
7	0

FIGURE 8.11
Stem-and-leaf diagram for the time required to complete a task on an assembly line: diagram suggests a nonnormal population.

when 55 is subtracted from each observation. From the stem-and-leaf diagram, it is easy to see that only three observations exceed 55. Thus the observed value of the test statistic Q_+ is 3. The P value of the test is found by computing the probability of seeing a value this small or smaller under the assumption that Q_+ is binomially distributed with $n = 15$ and $p = 1/2$. From Table I of the Appendix, $P = P[Q_+ \leq 3 | n = 15,\ p = 1/2] = .0176$. Since this P value is small, we reject H_0. We do have strong statistical evidence that the new procedure reduces the median time required to complete the task.

Since we assume that the underlying distribution is continuous, theoretically zero differences should not occur when conducting a sign test. However, as you might guess, sometimes zeros do occur in practice. These occur for various reasons but the primary problem is the lack of instruments capable of precise measurement of continuous phenomena such as time, length, speed, and volume. Treatment of zero differences has been considered extensively. Various recommendations as to how to treat these differences have resulted. These are our recommendations:

1. Assign to the zero differences the algebraic sign least conducive to the rejection of the null hypothesis. Thus, for a left-tailed test we would consider zero differences to be positive; for a right-tailed test they would be considered to be negative. In a two-tailed test we assign to zero differences the algebraic sign of the less frequently occurring difference. For example, if one observed 3 negative signs, 15 positive signs, and 6 zeros in running a two-tailed test, then the 6 zeros would all be treated as though they were negative. This procedure makes sense because a zero difference supports the null hypothesis that $M = M_0$. The suggested technique gives the null hypothesis the benefit of the doubt by making it harder to reject H_0.
2. If the number of zeros is small relative to the sample size n, discard these differences and reduce the sample size accordingly.

Occasionally a situation arises in which the differences $X_i - M_0$ are such that we can observe the algebraic sign of each difference but not its magnitude. In this case, the sign test is about the only choice available for testing location. Exercise 49 is an example of this type of problem. Usually, the actual numerical value of the differences can be obtained. Unfortunately, the sign test does not make use of this additional information. It treats a negative difference of $-.1$ in exactly the same way as it does a negative difference of -1000. For data in which the actual differences can be found, a second nonparametric test is available for testing for location. This test, the Wilcoxon signed-rank test, makes use of both the sign and magnitude of the observed differences $X_i - M_0$.

Wilcoxon Signed-Rank Test

In this test we assume that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a continuous distribution that is symmetric about an unknown median M. Consider the set of differences $X_i - M_0$, $i = 1, 2, 3, \ldots, n$ where M_0 is the hypothesized median of the distribution from which the sample is drawn. The null hypothesis to be tested is H_0: $M = M_0$ versus the usual alternatives H_1: $M > M_0$, H_1: $M < M_0$, or H_1: $M \neq M_0$. If H_0 is true, the differences $X_i - M_0$ are drawn from a distribution that is symmetric about zero. It is assumed that the differences are such that the magnitude as well as the algebraic sign of each can be obtained. To conduct the test, we form the set of n absolute differences $|X_i - M_0|$. These are then ranked from 1 to n in order of absolute magnitude with the smallest absolute difference receiving a rank of 1. These ranks, which we denote by $R_1, R_2, \ldots, R_n$, are then assigned the algebraic sign of the difference score that generated the rank. If H_0 is true, then each rank is just as likely to be assigned a positive sign as a negative one. Consider the statistics

$$W_+ = \sum_{\substack{\text{all}\\ \text{positive}\\ \text{ranks}}} R_i \qquad \text{and} \qquad |W_-| = \sum_{\substack{\text{all}\\ \text{negative}\\ \text{ranks}}} |R_i|$$

If H_0 is true, then we should expect W_+ and $|W_-|$ to be approximately equal. If $M > M_0$ then W_+ would tend to be too large and $|W_-|$ too small. Similarly, if $M < M_0$ we would expect the reverse to be true. Hence, we define our test statistic to be $W = \min(W_+, |W_-|)$. The exact distribution of W has been tabled for various values of the sample size n and significance level α. One such table is Table VIII of App. A. Using this table, we reject H_0 if the observed value of W is less than or equal to the stated critical value.

In practice, ties in the difference scores $X_i - M_0$ can occur. If ties occur, the values for each tied group should be given the midrank of the group. For example, suppose that we observe difference scores of $3, -3, 3$ which should occupy ranks 8, 9, and 10. We would assign each of the three values a rank of 9 and then assign the next largest difference score a rank of 11. Example 8.7.2 illustrates the idea.

Example 8.7.2. The melting point for a new lightweight material designed for use in automobile interiors is being investigated. It is known that due to impurities in the material the melting point is a random variable uniformly distributed over a small temperature interval. It is thought that the median melting point is less than 120°C. Do these data support this contention?

115.1	117.8	116.5	121.0
120.3	119.0	119.8	118.5

We are testing

$$H_0\colon M = 120$$

$$H_1\colon M < 120$$

We first subtract 120 from each observation and then find the absolute value of each difference.

x_i	115.1	120.3	117.8	119.0	116.5	119.8	121.0	118.5
$x_i - 120$	-4.9	.3	-2.2	-1.0	-3.5	$-.2$	1.0	-1.5
$\lvert x_i - 120\rvert$	4.9	.3	2.2	1.0	3.5	.2	1.0	1.5

We next rank these absolute differences from 1 to 8. Note that the value 1.0 occurs twice in what would normally be positions 3 and 4. We assign a rank of 3.5 to each of these values. The algebraic sign attached to each rank is the same as that of the difference that generated the rank.

$\lvert x_i - 120\rvert$	4.9	.3	2.2	1.0	3.5	.2	1.0	1.5
Rank	8	2	6	3.5	7	1	3.5	5
Signed rank	-8	2	-6	-3.5	-7	-1	3.5	-5

For these data

$$W_+ = \sum_{\substack{\text{all}\\ \text{positive}\\ \text{ranks}}} R_i = 2 + 3.5 = 5.5$$

$$|W_-| = \sum_{\substack{\text{all}\\ \text{negative}\\ \text{ranks}}} |R_i| = 8 + 6 + 3.5 + 7 + 1 + 5 = 30.5$$

Since the test is a left-tailed test, the test statistic is W_+. We reject H_0 if the observed value of this statistic is too small to have occurred by chance. From Table VIII of App. A with $n = 8$ we see that we can reject H_0 at the $\alpha = .05$ level (critical point = 6) but we are unable to reject at $\alpha = .025$ (critical point = 4). Thus the P value of the test lies between .025 and .05. Since this P value is fairly small, we reject H_0 and conclude that the median melting point of this material is below 120°C.

If the sample size n exceeds values given in Table VIII of App. A, a large sample normal approximation may be used. The statistic

$$\frac{W - E[W]}{\sqrt{\operatorname{Var} W}}$$

is approximately distributed as a standard normal random variable with

$$E[W] = \frac{n(n+1)}{4}$$

and

$$\operatorname{Var} W = \frac{n(n+1)(2n+1)}{24}$$

Exercise 53 illustrates this approximation procedure.

The Wilcoxon signed-rank test is almost as sensitive to departures from the null hypothesis as the normal theory T test even when the underlying distribution is normal. For other symmetric distributions, the signed-rank test is usually more powerful than the T test. Hence, this test should be considered a strong competitor to the T test for practical problems. This is particularly true for small samples where violations of the normal theory tests assumptions are of greatest concern.

Note that although a Wilcoxon signed-rank test does not assume normality it does assume symmetry. Procedures have been developed to test the validity of this assumption. One such test is given in [19].

CHAPTER SUMMARY

In this chapter we considered confidence interval estimation of the variance and standard deviation of a normal distribution. We also considered interval estimation of a mean when the population variance is unknown. This procedure entails the use of the Student-t or T distribution. We discussed this new continuous distribution in detail and saw that its properties are similar to those of the Z or standard normal distribution. In particular, we saw that for large sample sizes t points are well approximated by z points.

We next turned our attention to methods used in testing a statistical hypothesis. We found that we are always dealing with two hypotheses, the null hypothesis H_0 and its alternative H_1. The point of view of the researcher is stated as the alternative hypothesis. Thus, we hope that our data will allow us to reject H_0 thereby accepting H_1. We design our tests in such a way that we always know the probability of rejecting a true null hypothesis. We found that we are always subject to error when testing a hypothesis. If we reject a true null hypothesis, we commit a Type I error; if we fail to reject a false null hypothesis a Type II error is committed. Two methods were described for deciding whether or not to reject H_0. The first method is referred to as hypothesis testing. In conducting a hypothesis test, we preset α. This is done by setting up a rejection or critical region prior to data collection. We reject H_0 if the observed value of the test statistic falls into this critical region. The second method for deciding whether to reject H_0 is called significance testing. Here, no critical region is set prior to data gathering. Rather, we evaluate the test statistic and find the probability or P value of the test. The P value is the probability of observing a value of the test statistic as unusual or more unusual than that observed if the null value of the parameter θ is correct. Thus, the P value is the smallest value at which we could have preset α and still have been able to reject H_0. We reject H_0 if the P value is deemed to be small. There are advantages and disadvantages to each method. You should be familiar with both as they are both used extensively.

We considered in some detail what are commonly called "T tests." These are tests specifically designed to test a hypothesis on the mean of a normal distribution. We saw that these tests require that sampling be from a normal distribution and that this restriction is especially important for small samples. In Sec. 8.6 a method was given for testing for normality; in Sec. 8.7 we presented some nonparametric alternatives to the T test if the normality assumption appears to be invalid. Nonparametric tests are tests that make no assumption as to the family of distribution from which sampling is done.

Finally, we considered a method for testing a hypothesis on the variance or standard deviation of a normal distribution.

Many new terms and concepts were introduced in this chapter. Among them are the following:

Student-t distribution	Null hypothesis
Alternative hypothesis	Research hypothesis
Null value	Test statistic
Type I error	Type II error
α	β
Power	Size of test
Critical or rejection region	Level of significance
Significance test	Hypothesis test
Probability or P value	Critical level
Descriptive level of significance	Right-tailed test
Left-tailed test	Two-tailed test
Nonparametric test	Median

EXERCISES

Section 8.1

1. When programming from a terminal, one random variable of concern is the response time in seconds. These data are obtained for one particular installation:

1.48	1.26	1.52	1.56	1.48	1.46
1.30	1.28	1.43	1.43	1.55	1.57
1.51	1.53	1.68	1.37	1.47	1.61
1.49	1.43	1.64	1.51	1.60	1.65
1.60	1.64	1.51	1.51	1.53	1.74

(*a*) Construct a stem-and-leaf diagram. Does the assumption of normality appear reasonable?
(*b*) Find the unbiased point estimate for σ^2.
(*c*) Find a 95% confidence interval on σ^2.
(*d*) Find a 95% confidence interval on σ.
(*e*) Would you be surprised to hear the director of this installation claim that the standard deviation in response time is more than .2 second? Explain.

2. Highway engineers have found that the ability to see and read a sign at night depends in part on its "surround luminance." That is, it depends on the light intensity near the sign. These data are obtained on the surround luminance (in candela per square meter) of 30 randomly selected highway signs in a large metropolitan area. (Based on "Use of Retroreflectors in the Improvement of Nighttime Highway Visibility," H. Waltman, *Color*, pp. 247–251.)

10.9	1.7	9.5	2.9	9.1	3.2
9.1	7.4	13.3	13.1	6.6	13.7
1.5	6.3	7.4	9.9	13.6	17.3
3.6	4.9	13.1	7.8	10.3	10.3
9.6	5.7	2.6	15.1	2.9	16.2

(*a*) Find the sample variance for these data.

(*b*) Assume that the data are drawn from a normal distribution. Find a 90% confidence interval on the variance in the surround luminance in this area.

(*c*) Find a 90% confidence interval on the standard deviation in surround luminance.

(*d*) The normal probability rule (Sec. 4.5) implies that a normal random variable will lie within two standard deviations of its mean with probability .95. Use $\bar{X}$ and S to estimate the mean and standard deviation of the surround luminance in this area. Would it be unusual for the surround luminance for a randomly selected sign to exceed 18 cd/m^2? Explain.

3. X-ray microanalysis has become an invaluable method of analysis. With the electron microprobe both quantitative and qualitative measures can be taken and analyzed statistically. One method for analyzing crystals is called the two-voltage technique. These measurements are obtained on the percentage of potassium present in a commercial product which theoretically contains 26.6% potassium by weight: (Based on "Quantitative Electron Probe Analysis of Low Atomic Number Samples with Irregular Surfaces" by Klara Kiss, *Applied Spectroscopy*, February 1983, pp. 19–24.)

21.9	23.4	22.1	22.1	24.7	24.6
24.0	24.1	24.2	26.5	23.8	25.3
24.8	24.8	24.5	27.8	24.9	
27.2	25.1	25.5	23.7	26.5	
22.0	26.7	25.2	23.1	25.4	

(*a*) Check the reasonableness of the normality assumption by constructing a stem-and-leaf diagram for these data.

(*b*) Find the sample variance for these data.

(*c*) Find a 99% confidence interval for σ^2.

(*d*) Find a 99% confidence interval for σ. Note that this confidence interval is fairly long. Suggest a way to improve the interval estimate for σ based on these data. Try your suggestion to see if the new estimate is more informative than that given by the 99% confidence interval.

***4.** (*One-sided confidence interval on* σ^2.) Since variance is a measure of consistency, it is usually hoped that σ^2 will be small. For this reason, it is sometimes useful to construct what is called a "one-sided" confidence interval for σ^2. That is, we want to find an interval of the form $(0, L]$ where L is a statistic with the property that $P[\sigma^2 \le L] \doteq 1 - \alpha$. The derivation of such an interval is similar to that of the two-sided interval and is based on the diagram of Fig. 8.12. Use Fig. 8.12 and Theorem 8.1.1 to show that the upper bound for the one-sided confidence interval described is given by $L = (n - 1)S^2/\chi^2_{1-\alpha}$.

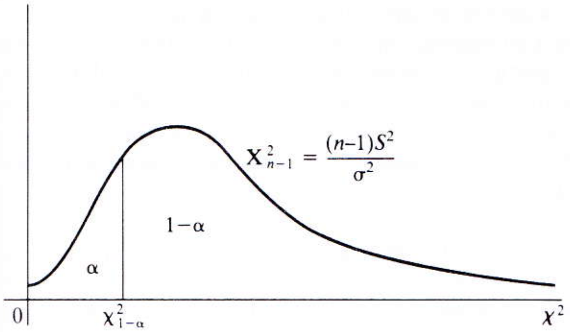


FIGURE 8.12
Partition of the X^2_{n-1} curve needed to derive a $100(1-\alpha)\%$ "one-sided" confidence interval on σ^2.

*5. Robotic technology is an area of rapid growth. It is reported that 315,000 industrial robots will be in use in American industry by the year 1995. One important feature of a robot is its accuracy. In a study of a particular robot used to apply adhesive to a specified location, these data are obtained on the error (in inches) in the placement of the adhesive: (Based on "Robotics Growth—No End in Sight", D. Hegland, *Production Engineering*, April 1983, pp. 46–51.)

.001	.002	.003	.002	.002
.007	.003	.004	.003	.006
.006	.003	.005	.004	.004
.001	.008	.001	.004	.003
.001	.003	.003	.005	.006

(*a*) Construct a stem-and-leaf diagram. Does the assumption that the placement error is normally distributed appear reasonable?
(*b*) Find the sample variance for these data.
(*c*) Use Exercise 4 to find 90% one-sided confidence intervals on σ^2 and σ.
(*d*) This robot is acceptable if its standard deviation does not exceed .005 inch. Does this criteria appear to be met? Explain.

*6. In Theorem 7.1.3 we showed that the sample variance is an unbiased estimator for σ^2 regardless of the distribution of the random variable X. If X is normal, this property is obtained more easily by making use of Theorem 8.1.1 and the properties of the chi-squared distribution given in Sec. 4.3. Use these results to show that for a normal random variable X, $E[S^2] = \sigma^2$ and $\operatorname{Var} S^2 = 2\sigma^4/(n-1)$.

*7. (*Normal Approximation to* χ^2.) Note that the chi-squared table lists values of γ from 1 to 30. Therefore it can be used for sample sizes from 2 to 31. For samples larger than this, chi-squared points can be approximated by the formula

$$\chi^2_r \doteq (1/2)\left[z_r + \sqrt{2\gamma - 1}\right]^2$$

For example, for a sample of size 50, the point $\chi^2_{.025}$ is given by

$$\begin{aligned}\chi^2_{.025} &\doteq (1/2)\left[z_{.025} + \sqrt{2\gamma - 1}\right]^2 \\ &= (1/2)\left[1.96 + \sqrt{2(49) - 1}\right]^2 \\ &\doteq 69.72\end{aligned}$$

(*a*) For a sample of size 100, approximate the points $\chi^2_{.05}$ and $\chi^2_{.95}$.

(*b*) Recent research indicates that heating and cooling commercial buildings with groundwater-source heat pumps is economically sound. The crucial random variable being studied is the water temperature. A sample of 150 wells in the state of California yields a sample standard deviation of 7.5 degrees Fahrenheit. Find a 95% confidence interval on the standard deviation in temperature of wells in California. (Based on data appearing in the *Consulting Engineer*, May 1983, p. 18.)

***8.** In pouring glass for use in automobile windshields uniformity of thickness is desirable to prevent distortion. Find a 95% one-sided confidence interval on the standard deviation in thickness if a sample of 100 windshields yields a sample standard deviation of 0.01 inches.

Section 8.2

9. Use the T table to find each of these points:

(*a*) $t_{.05}(\gamma = 8)$;
(*b*) $t_{.95}(\gamma = 8)$;
(*c*) $t_{.975}(\gamma = 12)$;
(*d*) $t_{.025}(\gamma = 12)$;
(*e*) $t_{.05}(\gamma = 121)$;
(*f*) $t_{.05}(\gamma = 150)$;
(*g*) Point t such that $P[-t \le T_{25} \le t] = .90$;
(*h*) Point t such that $P[-t \le T_{25} \le t] = .95$;
(*i*) Point t such that $P[T_{15} \ge t] = .05$;
(*j*) Point t such that $P[T_{20} \ge t] = .10$;
(*k*) Point t such that $P[T_{16} \le -t] = .05$;
(*l*) Point t such that $P[T_{30} \le -t] = .10$.

10. When the desired value of γ lies between two real numbered values listed in the T table there is a small problem! There are two solutions. You can use the value closest to γ or you can take a conservative approach and use the value that is a little smaller than the desired γ. The latter approach is conservative in that it results in confidence intervals that are a little longer than they need be. Thus, the actual confidence obtained is a little higher than the stated confidence level. Either approach is acceptable. When γ exceeds 120, we use the row labeled ∞. Find each of these points. Use the conservative value where applicable.

(*a*) $t_{.05}(\gamma = 50)$
(*b*) $t_{.025}(\gamma = 75)$
(*c*) $t_{.10}(\gamma = 200)$

11. Metal conduits or hollow pipes are used in electrical wiring. In testing one-inch pipes, these data are obtained on the outside diameter (in inches) of the pipe.

1.281	1.293	1.287	1.286
1.288	1.293	1.291	1.295
1.292	1.291	1.290	1.296
1.289	1.289	1.286	1.291
1.291	1.288	1.289	1.286

(*a*) For these data, $\Sigma_{i=1}^{20} x_i = 25.792$ and $\Sigma_{i=1}^{20} x_i^2 = 33.261596$. Use these values to find $\bar{x}$, s^2, and s for this sample.

(*b*) Assume that sampling is from a normal distribution. Find a 95% confidence interval on the mean outside diameter of pipes of this type.

(*c*) The makers of this type of pipe claim that the mean outside diameter is 1.29 inches. Does the confidence interval lead you to suspect this reported figure? Explain.

12. Lightweight hand-held laser range finders are now used by civil engineers in hydrographic surveys. In testing one brand of range finder these data are obtained on the error (in meters) made in locating an object at a distance of 500 meters: (*Civil Engineering*, February 1983, p. 52.)

−.10	−.02	.10	−.03	.09
.01	−.05	.05	−.06	.01
.03	.06	.02	−.07	.03

(*a*) For these data, $\sum_{i=1}^{15} x_i = .07$ and $\sum_{i=1}^{15} x_i^2 = .0489$. Use these values to find point estimates for the mean and standard deviation in the error made by the laser.

(*b*) Assume that these measurement errors are normally distributed. Find a 90% confidence interval on the mean measurement error.

(*c*) A competitor claims that this particular model on the average overestimates the distance by at least .05 meters. Is there reason to doubt the claim based on the observed data? Explain.

*(*d*) Based on the normal probability rule (Sec. 4.5) would you consider it unusual for a single measurement error to be in excess of .15 m? Explain.

13. One of the classic problems of operations research is the vehicle routing problem (VRP). This problem entails studying a system consisting of a given number of customers with known locations and demand for a commodity who are being supplied from a single depot by a number of vehicles with known capacity. The object of the study is to route the vehicles in such a way that the total distance traveled is minimized. The characteristics of a new algorithm, written in Pascal, are being investigated. These data are obtained on the cpu time required to solve the problem. (*European Journal of Operational Research*, April 1983, p. 388–393.)

2.0	1.4	3.5	2.3	3.2	3.6
.1	3.5	2.2	2.1	2.4	1.5
2.2	2.3	2.7	1.9	1.7	1.8
3.1	1.5	1.5	2.6	2.8	2.5
2.5	3.9	.8	1.8	3.3	3.7

(*a*) Estimate the mean and standard deviation in the time required to solve a problem via this algorithm.

(*b*) Find a 99% confidence interval on the mean time required to solve a problem.

(*c*) The best algorithm known to date for solving the problems studied is written in Algol and requires an average of 6.6 seconds cpu time. The solutions obtained are equivalent. Does the new algorithm appear to be more efficient than the old with respect to computing time? Explain.

14. To estimate the average number of pounds of copper recovered per ton of ore mined, a sample of 150 tons of ore is monitored. A sample mean of 11 pounds with a sample standard deviation of 3 pounds is obtained. Construct a 95% confidence interval on the mean number of pounds of copper recovered per ton of ore mined.

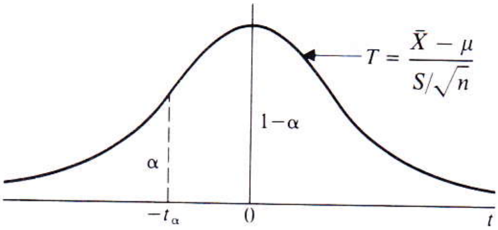


FIGURE 8.13

*15. (*One-sided confidence interval on μ.*) A "one-sided" confidence interval can be used to approximate the maximum or minimum value of a population mean.

(*a*) An interval of the form $(-\infty, L]$ such that $P[\mu \le L] \doteq 1 - \alpha$, allows us to place bounds on the maximum value of the population mean. Use Fig. 8.13 to show that $L = \bar{X} + t_\alpha S/\sqrt{n}$.

(*b*) An interval of the form $[L, \infty)$ allows us to place bounds on the minimum feasible value of the population mean. Show that $L = \bar{X} - t_\alpha S/\sqrt{n}$.

*16. These data are obtained on the total nitrogen concentration (in parts per million) of water drawn from a lake being considered for use as a source of drinking water for a locality:

.042	.023	.049	.036	.045	.025
.048	.035	.048	.043	.044	.055
.045	.052	.049	.028	.025	.039
.023	.045	.038	.035	.026	.059

Find a 95% one-sided confidence interval on the largest feasible value for μ. To be acceptable as a source of drinking water the mean nitrogen content must lie below .07 ppm. Does this lake appear to meet this criteria? Explain.

*17. (*Sample size required to estimate μ.*) Three factors determine the length of a confidence interval on μ. These are the confidence desired, the variability in the data, and the sample size. In an undesigned experiment it is possible that the resulting confidence interval is so long that it is almost useless. If σ is known or can be estimated from a small preliminary or "pilot" study, then it is possible to design an experiment in such a way that the resulting confidence interval will be short enough to be useful. This is done by selecting the sample size carefully.

(*a*) Let d denote the distance between $\bar{X}$, the center of the confidence interval and $\bar{X} + z_{\alpha/2}\sigma/\sqrt{n}$, the upper confidence bound. Thus, $d = z_{\alpha/2}\sigma/\sqrt{n}$. Note that the confidence interval itself is of length $2d$. Solve this equation for n to show that the sample size required to estimate μ to within d units with $100(1-\alpha)\%$ confidence is

$$n \doteq \frac{(z_{\alpha/2})^2\sigma^2}{d^2} \qquad \sigma \text{ known}$$

$$n \doteq \frac{(z_{\alpha/2})^2\hat{\sigma}^2}{d^2} \qquad \sigma \text{ unknown}$$

(*b*) Reading digital displays in bright light poses a problem. Engineers want to design a filter to maximize both the luminance (brightness) and the chrominance (color) contrast. To do so they intend to estimate the average number of footcandles in the cockpit of commercial airliners, where the filter will be used. A preliminary pilot study is run and an estimated standard deviation of 500 footcandles is obtained. How large a sample is needed to estimate μ to within 50 footcandles with 95% confidence? ("LED's Have a Place in the Sun," M. Christiansen, *Design News*, April 1983, pp. 41–52.)

(*c*) To determine whether or not the copper ore in a particular area is pure enough for open pit mining to be feasible, mining engineers must estimate the average grade of the ore. Past experience with this type of ore indicates that the grade ranges from 1 to 4% copper. The normal probability rule and Exercise 25, Chap. 6, imply that a rough estimate of σ is 1/4 of the range, or .75. How many test holes must be drilled to estimate μ to within .1% with 90% confidence?

*18. A study is being designed to estimate the mean time required to assemble a panel of microprocessor chips for use in color television sets. An estimate of this mean is needed in order to set reasonable quotas for assembly line workers. A small pilot study is conducted and these data are obtained on the assembly time in minutes.

1.0	1.5	2.2	3.0	2.7
2.0	2.4	2.6	2.3	1.7

(*a*) Based on these data estimate σ.

(*b*) How large a sample is required to estimate μ to within .2 minutes with 99% confidence?

Section 8.3

19. In 1969 in the United States, on average 8% of household waste was metal. Because of the increase in recycling efforts, it is hoped that this figure has been reduced. An experiment is run to verify this contention.

(*a*) Set up the appropriate null and alternative hypotheses for the experiment.

(*b*) Explain in a practical sense what has occurred if a Type I error has been committed.

(*c*) Explain in a practical sense what has occurred if a Type II error has been committed.

(*d*) Explain in a practical sense what it means to say that H_0 has been rejected at the $\alpha = .05$ level of significance.

20. The mean level of background radiation in the United States is .3 rem per year. It is feared that as a result of the increased use of radioactive materials, this figure has been increased.

(*a*) Set up the appropriate null and alternative hypotheses to document this claim.

(*b*) Explain in a practical sense the consequences of making a Type I and a Type II error.

21. As mentioned in Chap. 1, an important aspect of the engineering sciences is model building. Once a theoretical model is devised to explain a physical phenomena it must be tested to see that it yields results that are realistic. This testing is often done via computer simulation. In testing a model we are testing

$$H_0\text{: model is credible}$$

$$H_1\text{: model is not credible}$$

(a) Explain in a practical sense what has occurred if a Type I error is committed. The probability of committing this error is referred to as the "model builder's risk." Do you see why this language is appropriate?

(b) Explain in a practical sense what has occurred if a Type II error is committed. The probability of committing an error of this type is called the "model user's risk." Does this seem appropriate?

22. (*Power.*) The probability of making the *correct* decision of rejecting H_0 when H_1 is true is called the *power* of the test. Thus power $= 1 - \beta$. Where does power fit into the chart of Fig. 8.7? What is the power of the test of Example 8.3.4 when $p = .7$? When $p = .8$?

23. Suppose we want to test

$$H_0: p \leq .4$$
$$H_1: p > .4$$

based on a sample of size 15.

(a) Find the critical region for an $\alpha \doteq .05$ level test.

(b) If, when the data are gathered, $x = 11$, will H_0 be rejected?

24. Suppose we want to test

$$H_0: p \geq .7$$
$$H_1: p < .7$$

based on a sample of size 10.

(a) Find the critical region for an $\alpha \doteq .05$ level test.

(b) If, when the data are gathered, $x = 5$, will H_0 be rejected?

25. It is a common practice to subject long-life items to larger than usual stress so that failure data can be obtained in a short amount of test time. Such tests are called accelerated life tests. Equipment used in computing makes use of metal oxide semiconductors (MOS). It is thought that "oxide short circuits" account for a majority of the early failures found in MOS integrated circuits. To verify this contention, a high-voltage screen test is applied to a number of circuits and 15 early failures are observed. Let X denote the number of failures due to oxide short circuits. (Based on "Testing for MOS IC Failure Modes, D. Edwards, *IEEE Transactions on Reliability*, April 1982, pp. 9–17.)

(a) Set up the appropriate null and alternative hypotheses.

(b) If H_0 is true and $p = .5$, what is the expected number of failures due to oxide short circuits in the 15 trials?

(c) Let us agree to reject H_0 in favor of H_1 if X is 11 or more. In this way we are presetting α at what level?

(d) Find β if $p = .6$; if $p = .7$; if $p = .8$; if $p = .9$.

(e) Find the power of the test if $p = .6$; if $p = .7$; if $p = .8$; if $p = .9$.

(f) If, when the data are gathered, we observe 12 early failures that are due to oxide short circuits, will H_0 be rejected? What type error might be committed?

(g) If, when the data are gathered, we observe 10 early failures due to oxide short circuits, will H_0 be rejected? What type error might be committed?

26. Quality and reliability are becoming important aspects of computer hardware and software. Past experience shows that the probability of failure during the first 1000 hours of operation for 16-kbit dynamic RAMs produced by a United States firm is .2. It is hoped that new technology and stricter quality controls have reduced this failure rate. To verify this contention, 20 systems will be monitored for 1000 hours and the number of failures will be recorded. (Based on "Software Quality Improvement," Y. Mizuno, *Computer*, March 1983, pp. 66–72.)

(*a*) Set up the appropriate null and alternative hypotheses.

(*b*) Explain in a practical sense the consequences of making a Type I and a Type II error.

(*c*) If H_0 is true and $p = .2$, what is the expected number of failures during the first 1000 hours in the 20 trials?

(*d*) Let us agree to reject H_0 in favor of H_1 if the observed number of failures, X, is at most 1. In this way, we are presetting α at what level?

(*e*) Suppose that it is essential that the test be able to distinguish between a failure rate of .2 and a failure rate of .1. Find the probability that the test as designed will be unable to do so. That is, find β if $p = .1$. Find the power of the test if $p = .1$.

(*f*) The results of part (*e*) indicate that the test as designed cannot distinguish well between $p = .1$ and $p = .2$. Keeping the sample size fixed at $n = 20$, can you suggest a way to modify the test that will lower β and increase the power for detecting a failure rate of .1? Will α still be small enough to be acceptable? If not, can you suggest a way to redesign the experiment that will make both α and β low enough to be acceptable?

27. In Example 8.3.3, we test

$$H_0\colon p \le .5$$

$$H_1\colon p > .5 \qquad \text{(majority of automobiles in operation have misaimed headlights)}$$

at the $\alpha = .0577$ level by agreeing to reject H_0 if at least 14 of the 20 cars sampled have misaimed headlights. We claim that values of X that are too large to occur by chance when $p = .5$ are also too large to occur by chance when $p < .5$. That is, if these values are rare when $p = .5$ they are even more rare when $p < .5$. To help see that this is true, find $P[X \ge 14]$ when $p = .4; .3; .2; .1$. Are each of these probabilities less than .0577 as expected?

28. A sample of size 9 from a normal distribution with $\sigma^2 = 25$ is used to test

$$H_0\colon \mu = 20$$
$$H_1\colon \mu = 28$$

The test statistic used is the sample mean, $\bar{X}$. Let us agree to reject H_0 in favor of H_1 if the observed value of $\bar{X}$ is greater than 25.

(*a*) If H_0 is true, what is the distribution of $\bar{X}$?

(*b*) In the diagram of Fig. 8.14, shade the critical region for the test.

(*c*) Find α. Remember that α is computed under the assumption that H_0 is true.

(*d*) If H_1 is true, what is the distribution of $\bar{X}$?

(*e*) In the diagram of Fig. 8.14, shade the region whose area is β. Remember that β is computed under the assumption that H_1 is true.

(*f*) Find β.

(*g*) Find the power of the test.

(*h*) If the sample size is increased, the standard deviation of $\bar{X}$ will decrease. What is the geometric effect of this on the two curves of Fig. 8.14?

(*i*) If the sample size is increased but the critical point is not changed, what will be the effect on α and β?

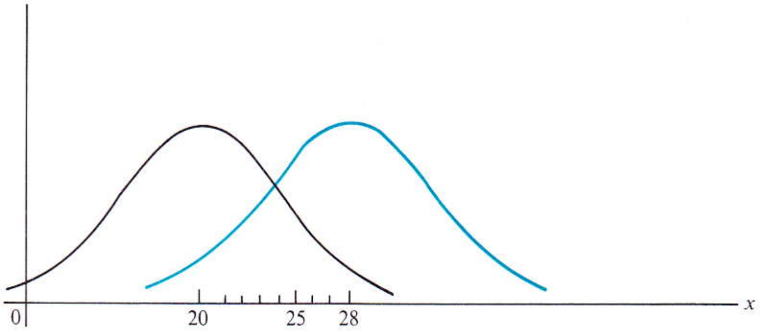


FIGURE 8.14

Section 8.4

29. Whenever a motorist encounters braking problems, especially an unpredictable pulling to one side, the villain is always held to be the brake pad. Trace elements, especially titanium, can combine with other elements to form minute particles of titanium carbonitride which alters the degree of friction between the pad and disc and leads to unequal wear. The percentage of titanium in a brake pad should not exceed 5%. A study is conducted to detect a situation in which the mean percentage of titanium in the brake pads being produced by a particular manufacturer exceeds 5%. (*Design Engineering*, February 1982, p. 24.)

(*a*) Set up the appropriate null and alternative hypotheses.

(*b*) Discuss the practical consequences of making a Type I and a Type II error.

(*c*) A sample of 100 brake pads yields a mean percentage of $\bar{x} = .051$. Assume that $\sigma = .008$. Find the P value for the test. Do you think that H_0 should be rejected? Explain. To what type error are you now subject?

30. The current particulate standard for diesel car emission is .6 g/mi. It is hoped that a new engine design has reduced the emissions to a level below this standard. (*Design Engineering*, February 1982, p. 13.)

(*a*) Set up the appropriate null and alternative hypotheses for confirming that the new engine has a mean emission level below the current standard.

(*b*) Discuss the practical consequences of making a Type I and a Type II error.

(*c*) A sample of 64 engines tested yields a mean emission level of $\bar{x} = .5$ g/mi. Assume that $\sigma = .4$. Find the P value of the test. Do you think that H_0 should be rejected? Explain. To what type error are you now subject?

31. It is thought that more than 15% of the furnaces used to produce steel in the United States are still open-hearth furnaces. To verify this contention a random sample of 40 furnaces is selected and examined.

(*a*) Set up the appropriate null and alternative hypotheses required to support the stated contention.

(*b*) When the data are gathered, it is found that 9 of the 40 furnaces inspected are open-hearth furnaces. Use the normal approximation to the binomial distribution (Sec. 4.6) to find the P value for the test. Do you think that H_0 should be rejected? Explain. To what type error are you now subject?

32. It is known that defective items will be produced even on automated assembly lines. A particular process typically produces 5% defectives. If the proportion of defectives exceeds 5% then the line must be shut down and adjusted.

(*a*) Set up the null and alternative hypotheses needed to detect a situation in which the proportion of defectives produced exceeds .05.

(*b*) Discuss the practical consequences of committing a Type I and a Type II error.

(*c*) A random sample of 100 items is selected and tested. Of these, 7 are found to be defective. Use the normal approximation to the binomial distribution to find the P value of the test. Do you think that H_0 should be rejected?

Section 8.5

33. Find the critical point(s) for conducting a hypothesis test on the mean with σ^2 unknown for

(*a*) a left-tailed test with $n = 25$; $\alpha = .05$

(*b*) a left-tailed test with $n = 150$; $\alpha = .10$

(*c*) a right-tailed test with $n = 20$; $\alpha = .025$

(*d*) a right-tailed test with $n = 16$; $\alpha = .01$

(*e*) a two-tailed test with $n = 20$; $\alpha = .10$

(*f*) a two-tailed test with $n = 30$; $\alpha = .05$

34. A new eight-bit microcomputer chip has been developed that can be reprogrammed without removal from the microcomputer. It is claimed that a byte of memory can be programmed in less than 14 seconds. (*Design News*, April 1983, p. 26.)

(*a*) Set up the null and alternative hypotheses needed to verify this claim.

(*b*) What is the critical point for an $\alpha = .05$ level test based on a sample of size 15?

(*c*) These data are obtained on X, the time required to reprogram a byte of memory:

11.6	14.7	12.9	13.3	13.2
13.1	14.2	15.1	12.5	15.3
13.3	13.4	13.0	13.8	12.3

Construct a stem-and-leaf diagram for these data. Does the normality assumption look reasonable?

(*d*) Test the null hypothesis. Can H_0 be rejected at the $\alpha = .05$ level? Interpret your result in a practical sense. To what type error are you now subject?

35. Ozone is a component of smog that can injure sensitive plants even at low levels. In 1979 a federal ozone standard of .12 ppm was set. It is thought that the ozone level in air currents over New England exceeds this level. To verify this contention air samples are obtained from 30 monitoring stations set up across the region. ("Air Pollution Stress and Energy Policy," F. Bormann, *Ambio*, vol. XI, 1982, pp. 188–194.)

(*a*) Set up the appropriate null and alternative hypotheses for verifying the contention.

(*b*) What is the critical point for an $\alpha = .01$ level test based on a sample of size 30?

(*c*) When the data are analyzed a sample mean of .135 and a sample standard deviation of .03 are obtained. Use these data to test H_0. Can H_0 be rejected at the $\alpha = .01$ level? What does this mean in a practical sense?

(*d*) What assumption are you making concerning the distribution of the random variable X, the ozone level in the air?

36. A model of Saudi Arabia's oil export strategy has been devised based on interviews with informed economists. The model is to be used to estimate the mean number of barrels of oil produced per day by this country. The usefulness of the model is to be partially checked by comparing the predicted mean for the year 1980 to its known value for that year, namely, 9.5 million barrels per day. ("Simulating Saudi Arabia's Oil Export Strategy," A. Picardi and A. Shorb, *Simulation*, January 1983, pp. 20–27.)

(*a*) Find the critical points for testing

$$H_0: \mu = 9.5$$

$$H_1: \mu \neq 9.5$$

at the $\alpha = .05$ level based on a sample of 50 simulations. (Use $\gamma = 40$ in Table VI of App. A.)

(*b*) For the data collected, $\bar{x} = 9.8$ and $s = 1.2$. Test H_0. Can H_0 be rejected at the $\alpha = .05$ level? Based on these data, is there evidence that the model is not adequate? To what type error are you now subject?

37. A new low-noise transistor for use in computing products is being developed. It is claimed that the mean noise level will be below the 2.5 dB level of products currently in use. (*Journal of Electronic Engineering*, March 1983, p. 17.)

(*a*) Set up the appropriate null and alternative hypotheses for verifying the claim.

(*b*) A sample of 16 transistors yield $\bar{x} = 1.8$ with $s = .8$. Find the P value for the test. Do you think that H_0 should be rejected? What assumption are you making concerning the distribution of the random variable X, the noise level of a transistor?

38. The Elbe River is important in the ecology of Central Europe as it drains much of this region. Due to increased industrialization it is feared that the mineral content in the soil is being depleted. This will be reflected in an increase in the level of certain minerals in the water of the Elbe. A study of the river conducted in 1982 indicated that the mean silicon level was 4.6 mg/litre. ("Natural and Anthropogenic Flux of Major Elements from Central Europe." T. Paces, *Ambio*, vol. XI, November 1982, pp. 206–208.)

(*a*) Set up the null and alternative hypotheses needed to gain evidence to support the contention that the mean silicon concentration in the river has increased.

(*b*) A sample of size 28 yields $\bar{x} = 5.2$ and $s = 1.6$. Find the P value for the test. Do you think that H_0 should be rejected?

39. Coal-handling maintenance is a very young technology. The emission standard for coal-burning plants is 4.8 lb. SO_2/million Btu/24-h average. In an attempt to get emissions below this level, engineers are experimenting with burning a blend of high- and low-sulfur coal. ("Upgrading and Maintaining Coal Handling," R. Rittenhouse, *Power Engineering*, March 1983, pp. 42–50.)

(*a*) Set up the null and alternative hypotheses needed to support the contention that the new mixture falls below the emission standard set by the government.

(*b*) Find the P value for the test if a sample of 200 readings yields a sample mean of 4.7 with a sample standard deviation of .5. Do you think that H_0 should be rejected? What does this mean in a practical sense?

40. Lasers are now used to detect structural movement in bridges and large buildings. These lasers must be extremely accurate. In laboratory testing of one such laser, measurements of the error made by the device are taken. The data obtained are used to test

$$H_0: \mu = 0$$
$$H_1: \mu \neq 0$$

A sample of 25 measurements yields $\bar{x} = .03$ mm over 100 m and $s = .1$. Find the P value for this two-tailed test. Do you think H_0 should be rejected? Interpret your result in a practical sense.

***41.** Confidence interval estimation on μ and hypothesis testing on μ, when σ^2 is unknown, are closely related. Both techniques make use of the T statistic $(\bar{X} - \mu)/(S/\sqrt{n})$. Show that if a given sample leads us to reject H_0: $\mu = \mu_0$ in favor of H_1: $\mu \neq \mu_0$ and the $\alpha = .05$ level, then a 95% confidence interval on μ based on the same data will *not* contain μ_0. *Hint:* Recall that H_0 is rejected if the observed value of the test statistic falls below the point $-t_{.025}$ or above the point $t_{.025}$.

***42.** (*Approximating sample sizes.*) In testing the hypothesis H_0: $\mu = \mu_0$, the experimenter can set α at any desired level. However, the value of β depends not only on the choice of α but also on the difference between μ_0 and the alternative value μ_1. The farther apart these values lie, the more likely it is that we will be able to distinguish them from one another. In designing an experiment, we want to pick a sample size that gives us a high probability of rejecting H_0 when there is a real practical difference between μ_0 and μ_1. That is, we want β to be small. Choosing the appropriate size for a T test is not easy. The problem is due to the fact that when H_0 is not true, our test statistic no longer follows a T distribution. Rather, it has what is called a noncentral T distribution. Fortunately, tables have been constructed using this distribution that allow us to determine the proper sample size for testing H_0: $\mu = \mu_0$ for various values of α, β, and Δ where $\Delta = |\mu_0 - \mu_1|/\sigma$ and σ is the standard deviation of X. Table VII of App. A is one such table. Its use is illustrated here.

Example. Let us test H_0: $\mu = 10$ versus H_1: $\mu > 10$ at the $\alpha = .05$ level. Assume that we want to be 90% sure of detecting a situation in which μ has gotten as large as 12. Assume also that a pilot study has been run and that $\hat{\sigma} = 4$. Here

$$\Delta \doteq |\mu_0 - \mu_1|/\hat{\sigma} = |10 - 12|/4 = .5$$
$$\alpha = .05 \qquad \text{and} \qquad \beta = .1$$

From Table VII of App. A we see that for a one-sided test with these characteristics we need a sample of size $n = 36$.

(*a*) A pilot study indicates that the standard deviation of a particular random variable X is 1.25. How large a sample is required to test

$$H_0: \mu = 20$$
$$H_1: \mu > 20$$

at the $\alpha = .05$ level and $\beta = .05$ level if it is important to be able to distinguish between $\mu = 20$ and $\mu = 21$?

(*b*) In Exercise 35, we tested

$$H_0: \mu = .12$$
$$H_1: \mu > .12$$

at the $\alpha = .01$ level based on a sample of size 30. From this study, we see that $\hat{\sigma} = .03$. Suppose that a mean ozone level of .14 is so serious that we must have a probability of .95 of detecting the situation. Approximately how large a sample is required?

(*c*) In Exercise 37, we tested

$$H_0: \mu = 2.5$$

$$H_1: \mu < 2.5$$

A sample of size 16 yielded $s = .8$. Assume that the new transistors are not financially worth marketing unless they reduce to mean noise level to at most 2 dB. Approximately how large a sample is needed to distinguish between a mean of 2.5 and a mean of 2.0 if $\alpha = .025$ and $\beta = .05$?

Section 8.6

43. A new process for producing small precision parts is being studied. The process consists of mixing fine metal powder with a plastic binder, injecting the mixture into a mold, and then removing the binder with a solvent. These data are obtained on parts which should have a one-inch diameter and whose standard deviation should not exceed .0025 inch.

1.0030	.9997	.9990	1.0054	.9991
1.0041	.9988	1.0026	1.0032	.9943
1.0021	1.0028	1.0002	.9984	.9999

For these data $\bar{x} = 1.0008$ and $s = .0028$.

(*a*) Use the Lilliefors graph of Fig. 8.15 to show that these data do not allow us to reject the normality assumption at the $\alpha = .05$ level.

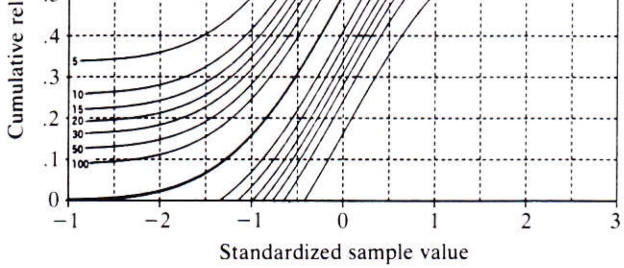


FIGURE 8.15
95% Lilliefors bounds for normal samples.

(*b*) Test

$$H_0: \mu = 1$$
$$H_1: \mu \neq 1$$

at the $\alpha = .05$ level.

(*c*) Test

$$H_0: \sigma = .0025$$
$$H_1: \sigma > .0025$$

at the $\alpha = .05$ level.

44. Indoor natatoriums or swimming pools are noted for their poor acoustical properties. The goal is to design a pool in such a way that the average time that it takes a low-frequency sound to die is at most 1.3 seconds with a standard deviation of at most .6 second. Computer simulations of a preliminary design are conducted to see whether these standards are exceeded. These data are obtained on the time required for a low-frequency sound to die. ("Acoustic Design in Natatoriums," R. Hughes and M. Johnson, *The Sound Engineering Magazine*, April 1983, pp. 34–36.)

1.8	3.7	5.0	5.3	6.1	.5
2.8	5.6	5.9	2.7	3.8	5.9
4.6	.3	2.5	1.3	4.4	4.6
5.3	4.3	3.9	2.1	2.3	7.1
6.6	7.9	3.6	2.7	3.3	3.3

For these data $\bar{x} = 3.97$ and $s = 1.89$.

(*a*) Use the Lilliefors graph of Fig. 8.16 to show that these data do not allow us to reject the normality assumption at the $\alpha = .01$ level.

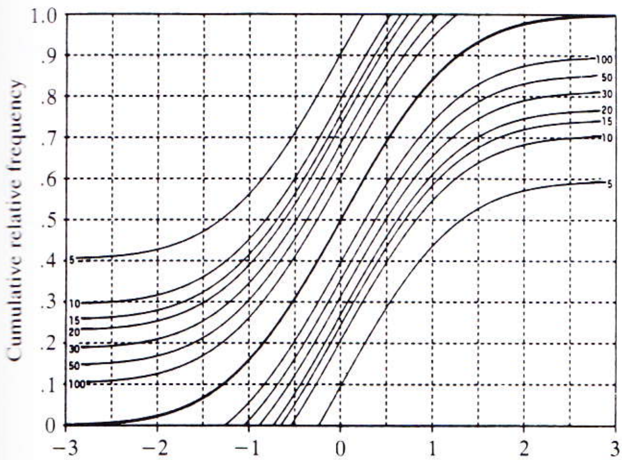


FIGURE 8.16
99% Lilliefors bounds for normal samples.

(*b*) Test

$$H_0: \mu = 1.3$$

$$H_1: \mu > 1.3$$

at the $\alpha = .01$ level.

(*c*) Test

$$H_0: \sigma = .6$$

$$H_1: \sigma > .6$$

at the $\alpha = .01$ level. Does it appear that the design specifications are being met?

45. Incompatibility is always a problem when working with computers. A new digital sampling frequency converter is being tested. It takes the sampling frequency from 30 to 52 kHz, word lengths of 14 to 18 bits, and arbitrary formats and converts it to the output sampling frequency. The conversion error is thought to have a standard deviation of less than 150 picoseconds. These data are obtained on the sampling error made in 20 tests of the device ("The Compatability Solution," K. Pohlmann, *The Sound Engineering Magazine*, April 1983, pp. 12–14.)

133.2	−11.5	−126.1	17.9	139.4
−81.7	314.8	147.1	−70.4	104.3
56.9	44.4	1.9	−4.7	96.1
−57.3	−43.8	−95.5	−1.2	9.9

For these data $\bar{x} = 28.69$ and $s = 104.93$.

(*a*) Use the Lilliefors graph of Fig. 8.17 to show that these data do not allow us to reject the normality assumption at the $\alpha = .1$ level.

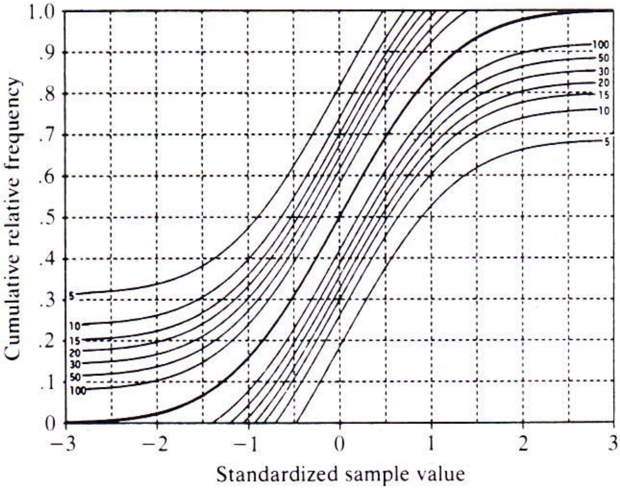


FIGURE 8.17
90% Lilliefors bounds for normal samples.

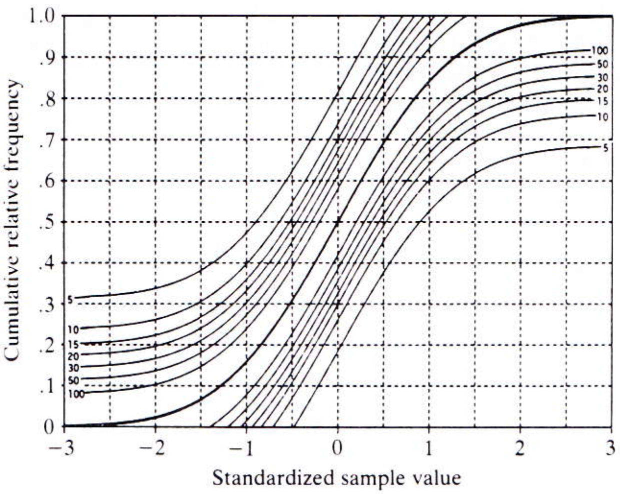


FIGURE 8.18
90% Lilliefors bounds for normal samples.

(b) Test

$$H_0: \mu = 0$$
$$H_1: \mu \neq 0$$

at the $\alpha = .1$ level.

(c) Test

$$H_0: \sigma = 150$$
$$H_1: \sigma < 150$$

at the $\alpha = .1$ level. Does the converter appear to be as accurate as claimed?

*46. Consider the data of Example 6.2.1 concerning the life span of a lithium battery. Based on a stem-and-leaf diagram, we have reason to suspect that these data are not drawn from a normal distribution. Use the Lilliefors graph of Fig. 8.18 to test for normality at the $\alpha = .1$ level. Are our suspicions justified?

Section 8.7

47. In each case, use the sign test to decide whether H_0: $M = M_0$ will be rejected in favor of the stated alternative at the $\alpha = .05$ level based on the data given. Do not discard zeros.

(a) H_1: $M > M_0$; $n = 15$, $Q_+ = 13$, no zeros
(b) H_1: $M > M_0$; $n = 20$, $Q_+ = 15$, no zeros
(c) H_1: $M > M_0$; $n = 20$, $Q_+ = 15$, three zeros
(d) H_1: $M < M_0$; $n = 10$, $Q_+ = 1$, no zeros
(e) H_1: $M < M_0$; $n = 10$, $Q_+ = 1$, one zero
(f) H_1: $M \neq M_0$; $n = 15$, $Q_+ = 2$, no zeros
(g) H_1: $M \neq M_0$; $n = 15$, $Q_+ = 2$, one zero

In each case above, what is the P value of the test?

48. Engineers are designing the safety devices for use in a new amusement-park ride. They think that the median height of patrons of rides of this sort exceeds 68 inches. Based on the sign test, do these data support this contention? Support your answer by finding the P value of the conservative sign test.

Height in inches

65	73	72	71	68
74	74	66	68	69
70	66	72	67	73
69	70	73	70	74

49. Even with careful workmanship, digital scales may need some adjustment before being put into use. Unless there are systematic errors being made, the apparent zero of the scales before adjustment should fluctuate about true zero. That is, some scales should weight a little heavy whereas others should give readings that are a little light. Ten such scales are randomly selected and tested. These data are obtained on the accuracy of the zero reading:

heavy	light	heavy	heavy	heavy
light	light	light	heavy	heavy

Based on these data can we reject H_0: $M = 0$ in favor of H_1: $M > 0$ at the $\alpha = .05$ level?

50. In Example 8.7.2 we were able to reject

$$H_0: M = 120$$

$$H_1: M < 120$$

at the $\alpha = .05$ level. If we had used the sign test, which ignores the magnitude of the difference scores, could we have rejected H_0 at the $\alpha = .05$ level? Explain by finding $P[Q_+ \leq 2 | n = 8 \text{ and } p = 1/2]$.

51. An experiment for treating tar sand wastewater was conducted to determine if a new treatment process removed more total organic carbon than a standard treatment process which is known to remove a median of 40 mg/litre in a fixed detention time. Under the same experimental conditions, the new process was replicated 10 times yielding total organic carbon amounts removed of 38.8, 53.6, 39.0, 51.6, 40.1, 46.9, 40.9, 44.9, 41.0, and 43.2.

(a) What is $E[W]$?

(b) Using the signed-rank test, is there evidence that the new process removes significantly more total organic carbon than the standard process at the .05 level?

52. In an attempt to determine how many consultants are needed to answer questions of users at a computer center, these data are collected on X, the time in minutes required to answer a telephone inquiry:

1.5	1.0	5.0	1.9	3.0
1.3	2.1	1.7	6.5	4.2
6.3	5.6	5.1	2.5	6.9

(*a*) What is $E[W]$?

(*b*) Based on the signed-rank test, can we conclude that the median time required is less than 5 minutes? Explain, based on the P value of your test. (A zero score should be given the lowest rank and should be assigned the algebraic sign least conducive to rejecting the null hypothesis.)

*53. A study of the expansion joints used in bridge beds is conducted. It is thought that these joints are expanding more than they were designed to expand thus creating cracks in the pavement near the joint. The median design expansion at 95°F is two inches. Laboratory tests of 100 such joints are conducted at this temperature.

(*a*) What is $E[W]$?

(*b*) What is $\text{Var}[W]$?

(*c*) Set up the appropriate null and alternative hypotheses.

(*d*) If $|W_-| = 1600$, can H_0 be rejected? Explain, based on the P value of the test.

REVIEW EXERCISES

54. A consumer group wants to estimate the mean cost of the base system for a personal computer with certain specifications. It is thought that these computers range in price from \$2390 to \$4000.

(*a*) How large a sample should be taken to estimate μ to within \$100 with 90% confidence?

(*b*) A random sample of size 50 yields these data: (data in thousands of dollars)

2.43	2.86	2.74	2.75	2.69	2.64	2.91
2.89	3.18	3.00	3.21	3.07	3.72	3.24
3.17	3.57	3.37	3.56	3.30	2.32	3.09
2.99	3.20	3.25	3.70	3.45	2.82	2.88
2.71	3.25	2.86	2.93	3.45	3.11	3.86
2.96	3.00	2.88	3.19	3.56	3.21	3.33
3.39	3.14	2.90	3.49	3.02	3.56	2.87
2.32						

Construct a stem-and-leaf chart for these data. Use the digits 2 and 3 as stems five times each. Graph numbers beginning 2.0 and 2.1 on the first stem, those beginning 2.2 and 2.3 on the second stem, and so forth. Does the stem-and-leaf chart lead you to suspect that these data are not drawn from a distribution that is at least approximately normal?

(*c*) Find unbiased estimates for μ and σ^2 based on these data. Estimate σ. Is the estimate for σ unbiased?

(*d*) Find 90% confidence intervals on σ^2 and σ.

(*e*) Find a 90% confidence interval on μ.

55. Researchers are experimenting with a new compound used to bond Teflon to steel. The compounds currently in use require an average drying time of three minutes. It is thought that the new compound dries in a shorter length of time.

(*a*) Set up the null and alternative hypotheses needed to support the claim that the new compound dries faster than those currently in use.

(*b*) Discuss the practical consequences of making a Type I error; a Type II error.
(*c*) A pilot study shows that $\hat{\sigma} = .5$. Suppose that the new product is worth marketing if the average drying time can be shown to be 2.5 minutes or less. How large a sample is required to detect this situation with probability .95 with α set at .05?
(*d*) When the experiment is conducted, these data are obtained:

1.4	2.1	2.8	.9
2.4	1.7	3.7	2.7
2.6	1.9	2.8	2.8
2.2	2.2	3.4	1.9

Test the null hypothesis of part (*a*) at the $\alpha = .05$ level. Would you suggest marketing this new product?

56. It is thought that a majority of the procedures used in a statistical computer package run in less than .1 second. To verify this contention, a random sample of 20 programs which entail exactly one procedure is to be examined.
(*a*) Set up the appropriate null and alternative hypotheses needed to verify the claim.
(*b*) Let X denote the number of programs in which the procedure used runs in less than .1 second. Find the critical region for an $\alpha \doteq .025$ level test.
(*c*) When the test is conducted, 14 programs are found in which the procedure used runs in less than .1 second. Will H_0 be rejected? To what type error are you now subject?
(*d*) Find β if $p = .6$; if $p = .7$; if $p = .8$; if $p = .9$.
(*e*) Find the power of the test if $p = .6$; if $p = .7$; if $p = .8$; if $p = .9$.

57. Nickel powders are used in coatings used to shield electronic equipment from electromagnetic interference. It is thought that the mean size of the individual nickel particles in one such coating is less than three micrometers. Do these data support this contention? Explain, based on the P value of the appropriate test.

3.26	3.07	2.46	1.76
1.89	2.95	3.35	3.82
2.42	1.39	1.56	2.42
2.03	3.06	1.79	2.96

58. We want to test

$$H_0: \mu = 5$$
$$H_1: \mu > 5$$

based on a random sample of size 25. The sample standard deviation is 2 and the observed value of the sample mean is 5.5. What is the P value for the test?

COMPUTING SUPPLEMENT

II One-Sample T Tests and Confidence Intervals

If you have not read the computing supplement at the end of Chap. 6, do so now. We present here an SAS program that will generate the statistics needed to construct confidence intervals on μ, σ^2, or σ. It will also test a null hypothesis on the value of μ and print its P value. It utilizes PROC MEANS introduced in Chap. 6. Data used are the data of Exercise 34.

Statement	*Function*
DATA CHIP;	names data set
INPUT TIME;	names variable
STIME = TIME − 14;	creates a new random variable whose hypothesized mean is 0; the constant subtracted is the null value of μ
CARDS;	signals beginning of data lines
11.6	
13.1	
13.3	data (one observation per line)
⋮	
12.3	
;	signals end of data
PROC MEANS MEAN VAR STD STDERR MAXDEC = 3; VAR TIME;	asks for needed sample statistics for the variable TIME
TITLE1 MAKING INFERENCES; TITLE2 ON THE MEAN;	titles the output
PROC MEANS T PRT MAXDEC = 3; VAR STIME;	tests the null hypothesis H_0: $\mu = 14$

The printout of this program follows. Note that

① gives the observed value of $\overline{X}$;
② gives the observed value of S^2
③ gives the observed value of S;
④ gives the observed value of $S/\sqrt{n}$;
⑤ gives the observed value of the test statistic $(\overline{X} - 14)/(S/\sqrt{n})$ used to test H_0: $\mu = 14$;
⑥ gives the P value for a two-tailed test; the P value for a one-tailed test is half of this value or .0279.

MAKING INFERENCES
ON THE MEAN

VARIABLE	MEAN	VARIANCE	STANDARD DEVIATION	STD ERROR OF MEAN
TIME	13.447	1.056	1.027	0.265
	①	②	③	④

MAKING INFERENCES ON THE MEAN		
VARIABLE	T	PR > !T!
STIME	−2.09	0.0558
	⑤	⑥

III Testing for Normality

The procedure PROC UNIVARIATE generates a wide variety of summary statistics including those generated by PROC MEANS. In addition it can be used to construct a stem-and-leaf diagram for a data set and to test for normality. The test used is not the Lilliefors test. Rather it uses the Shapiro-Wilk W Statistic or a modified Kolmogorov-Smirnov statistic. In either case the hypothesis of normality is rejected for small values of W. We illustrate its use by testing the data of Example 8.6.1 for normality.

Statement	*Function*
DATA AUTO;	names the data set
INPUT DISPLACE;	names the variable
CARDS;	signals beginning of data lines
6.2	
4.6	
4.1	data (one observation per line)
⋮	
3.2	
;	signals end of data
PROC UNIVARIATE PLOT NORMAL;	asks for a stem and leaf plot and for a test for normality
TITLE1 IS THE DISTRIBUTION;	titles the output
TITLE2 NORMAL?;	

The printout of this program follows. Note that

① gives a stem-and-leaf plot for the data;
② gives the value of W, the statistic used to test for normality;
③ gives the P value for the test for normality. Since $P = .381$ is large, we are unable to reject the null hypothesis that the data are drawn from a normal distribution.

IS THE DISTRIBUTION
NORMAL?

UNIVARIATE

VARIABLE = DISPLACE

MOMENTS				QUANTILES(DEF = 4)			
N	20	SUM WGTS	20	100% MAX	6.2	99%	6.2
MEAN	3.385	SUM	67.7	75% Q3	4.35	95%	6.135
STD DEV	1.40947	VARIANCE	1.98661	50% MED	3.7	90%	4.89
SKEWNESS	−0.143406	KURTOSIS	−0.670715	25% Q1	2.05	10%	1.31
USS	266.91	CSS	37.7455	0% MIN	1.1	5%	1.11
CV	41.6387	STD MEAN	0.315167			1%	1.1
T: MEAN = 0	10.7403	PROB > !T!	0.0001	RANGE	5.1		
SGN RANK	105	PROB > !S!	0.0001	Q3 − Q1	2.3		
NUMˆ = 0	20			MODE	3.7		
W: NORMAL	0.947402	PROB < W	0.381				
	2		3				

EXTREMES

LOWEST	HIGHEST
1.1	4.4
1.3	4.6
1.4	4.8
1.5	4.9
1.9	6.2

```
      STEM  LEAF                          #      BOXPLOT
         6  2                             1
         5                                          !
 1       4  1224689                       7         !
         3  25779                         5     +-----+
         2  56                            2     *--+--*
         1  13459                         5     +-----+
            ----+----+----+----+                   !
```

NORMAL PROBABILITY PLOT

```
6.5+                                                              *
   !                                                          +
   !                                           * * * * *   *   *  *
   !                                    * * * * * * * ++
   !                              +  + *   * +
1.5+               *       *    *   *   *
   +----+----+----+----+----+----+----+----+----+----+----+
        -2        -1        +0        +1        +2
```